THIS WEEK

Nailing fingerprints in the stars

Laboratory–based experiments are sorely needed to complement the rapidly proliferating spectral data originating from observations by the latest space telescopes.

What are stars made of? After astronomers detected a bright-yellow, unknown spectral line in sunlight in 1868, they named the new element helium after the Greek Sun god Helios. But it was some 30 years before physicists on Earth managed to detect — and so confirm the discovery of — helium in a laboratory.

It is a pattern that has been repeated many times since: the indirect detection of elements and molecules through spectral signatures in space has come ahead of detailed study on the ground. Lab spectroscopy has long lagged behind telescope observations, but it is striking just how wide the gap has now grown.

A cutting-edge infrared spectrograph, for example, installed in 2011 on the Sloan Digital Sky Survey (SDSS) telescope in Sunspot, New Mexico, records the spectra of 1,800 stars per night, most located in the bulge of the Milky Way galaxy, where dust prevents visible wavelengths of light from reaching Earth. The result is the detection of thousands of unidentified spectral lines — dips or peaks of electromagnetic waves at specific energies, caused by absorption of the light by gas on the way to Earth or emission by gas on stars.

Some physicists are now pointing out the irony that multimillion-dollar projects such as the SDSS are producing data that cannot be analysed because of a failure to support much cheaper lab work on the ground. They have a point, and support for lab-based research that can decipher such spectra should be increased. A good rule of thumb is that agencies funding telescope projects that are doing cutting-edge spectroscopy should spend a small fraction, maybe a few per cent, of the money on associated lab spectroscopy.

Lab-based measurements are less glamorous, but big questions about the evolution of galaxies will be solved by understanding small but important details about the physics and chemistry of millions of stars as revealed by spectra. For example, spectra could give clues to whether stars in the Galactic bulge formed there or migrated to it later. Spectra can also shed light on the amount of dark matter near a star, by reveal-ing information about the star's motion, which shifts its spectral lines.

A good example of the benefits of such work comes from a Novem-ber paper in *The Astrophysical Journal* by atomic physicists at Imperial College London, the National Institute of Standards and Technology in Gaithersburg, Maryland, and the Astrophysical Institute of the Canary Islands in Tenerife, Spain (M. P. Ruffoni *et al. Astrophys. J.* **779,** 17; 2013). They report 28 probabilities of electron transitions between sets of energy levels for the element iron. These can now be used in combination with spectra to estimate the abundance of iron in stars in the Galactic bulge — a step towards determining their ages and where they formed. None had previously been measured in the laboratory.

Such research is necessary because, to identify and quantify ele-ments in space from spectra, astronomers must know the probability that electrons in the elements' atoms will move between energy levels. For light elements with few electrons, such as hydrogen and helium, the probabilities of transitions can be calculated using the rules of quantum mechanics. But heavier elements have many electrons that can participate in transitions — iron has 26, making the probabilities of possible transitions between levels too complex to calculate accu-rately. Measuring emissions in the lab is the only alternative. Physicists can use tunable lasers to excite electrons into more levels and measure further transitions. This information can then feed back to the astro-nomical observations. Extra funds would significantly improve this capacity, giving better access to powerful lasers and detectors.

Even as experimentalists face challenges taking lab spectra, there is an astronomical spectroscopy boom. Aside from the infrared instru-ment taking data on the US$55-million SDSS, astronomers are planning to build gigantic 30–50-metre telescopes, such as the €1-billion (US$1.3-billion) European Extremely Large Telescope, to be based near Cerro Paranal, Chile, which will take hundreds of thousands of stellar spectra. Furthermore, NASA's planned $8.8-billion James Webb Space Telescope, which like the Sloan instrument uses cutting-edge mercury cadmium telluride infrared detectors, will look at stars and, it is hoped, at the atmospheres of planets outside the Solar System. Although the spectra can be used to estimate the amounts of different elements in the atmospheres of stars or planets, a particular area of inter-est is in identifying molecules, which also emit characteristic spectral lines when they transition between different states.

> *"Measuring emissions in the lab is the only alternative."*

Other lab-based experiments might even solve one of the longest-standing questions in astronomy: the origin of the diffuse interstellar bands — dips in the spectra of stars caused by diffuse matter spread between the stars and Earth. They are thought to be due to unstable hydrocarbon radicals, the exact mix of which has yet to be made in the laboratory, and they have puzzled astronomers for almost 100 years. How long do researchers want to wait? ∎

The DIY dilemma

Misconceptions about do–it–yourself biology mean that opportunities are being missed.

The do-it-yourself-biology movement has an image problem. More commonly called DIYbio, it tends to conjure up pictures of T-shirt-clad misfits marshalling limited scientific skill in their basements as they try to make cool-but-fringe things such as glow-in-the-dark plants. Policy-makers take an opposite view: instead of wayward amateurs, they see twisted experts hellbent on harm, engineering pathogens in their garages to unleash upon the world. A

survey of DIY biologists released on 19 November by the Woodrow Wilson International Center for Scholars in Washington DC reveals, unsurprisingly, that neither caricature is accurate, and that the DIYbio movement is more nuanced than it would seem to those looking in from the outside (see go.nature.com/nj9xk6).

The movement is made up of enthusiasts with a range of backgrounds and interests in biology, who work in wet-lab spaces not affiliated with traditional science centres such as universities. The survey found that 92% of DIY biologists work at least some of the time in communal spaces rather than in their garages or basements; that they are mostly young (36% under 35, 78% under 45); that they are more educated than the general population; and that many are still learning the basics of biotechnology. Only 6% of people surveyed said their experiments were of the kind that would require the safety conditions for work that might cause human disease.

It is interesting to note that 28% of people who responded to the survey said that they already do some or all of their work in academic, corporate or government labs, and that 19% have obtained a doctorate-level degree. So at least some DIY biologists are peers of — or indeed themselves — readers of this journal, and are within the mainstream scientific community.

This undercuts the notion that all DIY biologists are inexperienced if enthusiastic amateurs. And the report argues that this expertise and access to sophisticated lab facilities mean that the DIY community has the potential to generate products that will benefit society. As a result, it recommends that the US government should fund networks of community lab spaces.

Examples of the positive impact DIYbio can have already exist: its practitioners have produced a cheap alternative to commercial machines for the polymerase chain reaction, and they have come up with an inexpensive diagnostics device for malaria. Yet so far, the projects that have garnered the most attention have been essentially frivolous, such as the project to create a glowing plant, which collected US\$500,000 in public crowdsourced funds last year — ten times as much as the malaria tool earned in seed funding.

This highlights the key problem. There is no government granting agency judging which DIY project is worthwhile, so DIY biologists can do what they like, as long as it's legal. Although this is an intrinsic part of the thrill of being in the movement, it is also a factor that keeps legitimate funders away, and some community labs are threatened with closure as a result. Governments would gain much by supporting the DIYbio movement; it would give them more access to and potentially more control over the work that goes on in labs that they fund.

> "DIY biologists can do what they like, as long as it's legal."

But the report also notes that most DIY biologists do not favour government regulation, now or in the future. Governments, of course, cannot become more involved in supporting this movement without taking a more proactive role towards regulation. Is this apparent impasse permanent? Perhaps not. The report notes that a sizeable minority — 43% — of DIY biologists do favour some kind of regulation in the future, and this may grow as the movement matures.

The report's authors anticipate such a change. They suggest benchmarks and timelines to address regulation — a time in the future, for instance, when people outside companies and sophisticated labs will be able to synthesize long stretches of DNA. Still, rather than risk being overrun by events, the DIY-biology community and regulators should start to talk about how to anticipate such developments, rather than merely respond to them.

The security and stability of government funds would safeguard the future of the DIYbio movement; the issue is whether the movement would accept the trade-offs that such stability would bring. If you are reading, then do please tell. ∎

Enemy of the good

Universities need to counter pressures that undermine support for younger researchers.

Who are the outstanding mentors of young researchers? Since 2005, *Nature* has awarded an annual prize for scientific mentoring, rotating through a variety of countries. Over the years it has become clear that, regardless of the country and scientific discipline, there are some consistent key characteristics of lab heads that bode particularly well for young scientists under their leadership. Outstanding mentors tend to have a thorough command of their research field. They are highly accessible to the members of their lab. They can relate to individuals in a way that is specific to each person's characteristics. And they know how to balance support with the nurturing of independent creativity, problem-solving, integrity and initiative (see *Nature* **447,** 791–797; 2007).

This year's winners are no exception. The competition was held in Italy, and the awards went to neurobiologist Michela Matteoli, theoretical physicist Giorgio Parisi and chemist Vincenzo Balzani (see pages 443 and 559). All received glowing testimonials from their past trainees. For example, the success of one mentor was ascribed to "complete emotional and scientific investment" in mentees, who in turn "dedicate themselves to work at their best to pay back that faith".

That degree of mentoring commitment is unusual. All too often one meets young researchers who, despite working in prestigious institutions, have had no such experience. Yes, the 'sink or swim' approach can breed resilience, but proper mentoring can safeguard scientific integrity in the full sense of the word. It enables young researchers to develop a critical approach to their own ideas and data, and to maintain professionalism by using robust techniques and analyses. Mentoring also helps to engender a culture of transparency in allowing others access to raw data, gives a sense that one's leader has one's interests at heart, and can moderate the pressure to publish. Universities have a duty to ensure that this culture prevails, not least to ensure that public and private money is not squandered on sloppy, amateurish research.

But especially now, the pressures on young lab leaders are huge. Encounters with early-career principal investigators all too often indicate how narrow their focus must be to survive. They might be adding to those pressures because of hyper-competitiveness or anticipated demands from university and funding-agency committees. Typically, principal investigators are well-intentioned towards their younger colleagues, but feel an obligation to produce strong results in the first few years of their labs, to get funding or tenure. They may often feel that they do not have enough time to invest in mentoring their teams. Or they may well judge that they simply cannot tolerate people in their labs who are underperforming.

Such a lack of attention to nurturing individuals could exacerbate another damaging trend. With more people seeking alternative careers during their PhDs because of the ever tougher prospects in academia, those graduate students might lose motivation to go the extra mile to fulfil their research potential. And yet the principal investigator needs the papers generated by the students' work to get tenure.

These problems can be addressed in two ways: from the bottom up, by a sheer determination of younger lab heads to be responsible leaders; and more importantly, from the top down, by heads of universities and departments providing incentives for great leadership. Such heads should look at the winners of the *Nature* mentoring awards and ask: 'Does my institution cultivate such behaviour or hinder it?' ∎

↻ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv

SARAH HAY

Football fever could be a dose of dengue

Fans at next year's World Cup in Brazil may be exposed to a nasty and incurable tropical disease, warns **Simon Hay***.*

The twentieth FIFA World Cup will take place in Brazil in June and July next year. This football tournament is expected to sell more than 3 million tickets and attract more than half a million international fans. But those who attend will have more to worry about than the fitness of their top goalscorers: dengue fever could be a significant problem in some of the tournament locations, and preventive measures are needed. Dengue is a persistent threat to Brazilians, as it is to billions of people throughout the tropics. It is much less familiar to others, such as Europeans. This means that FIFA, the Brazilian authorities and the World Cup sponsors must use their influence and experience to communicate the risk and what protective measures fans should take.

Next week sees the draw for the group-stage matches, which will help fans to plan their trips. One thing we know already is that the dengue risk will be close to its peak when matches are played in three of the host cities: Fortaleza, Natal and Salvador, all in the northeast of the country. Much could be done by the authorities there to reduce dengue risk in the run-up to the tournament.

Dengue is a viral infection that can produce a severe fever and symptoms that may require hospitalization. It is transmitted to (and between) humans by urban-adapted, day-biting *Aedes* mosquitoes and is therefore a particular problem in towns and cities. To explore this risk, my colleagues and I assessed the potential levels of exposure by examining distribution maps for dengue in Brazil and records of its seasonal variation at key sites (full details, credits and maps are on my website at go.nature.com/8g1io5).

Like the weather, it is impossible to forecast the precise situation with regard to dengue in Brazil in 2014. We can, however, make informed guesses on the basis of averaged records of dengue in previous years. For the areas around nine of the World Cup stadiums, these records show that the main dengue season will have passed before the World Cup is held in June and July. Unfortunately, the risk remains high during these months in the northeast.

The Brazilian authorities should implement aggressive vector control in April and May, particularly around the northern stadiums, to decrease the number of dengue-transmitting mosquitoes. They can target adult *Aedes* mosquitoes through fogging (the use of aerosol formulations of insecticides that disperse efficiently) and can interrupt breeding by clearing sites at which the mosquitoes lay their eggs — water collected in discarded rubbish, for example. Although control efforts have failed to stem the worldwide increasing incidence of dengue and the expansion of its endemic range,

considerable local, albeit transient, reductions in mosquito populations have been achieved in some places, including Singapore.

There are no vaccines or drugs against dengue, but an individual will never contract dengue if they do not get bitten by an infected mosquito in the first place. So avoiding mosquito bites is the best precaution. Select accommodation with screened windows and doors and air conditioning; use insecticides indoors; wear clothing that covers the arms and legs, especially during early morning and late afternoon, when the chance of being bitten is greatest; and apply insect repellent to clothing and exposed skin.

The mass gatherings and predictable movement of fans should be a help to campaigns promoting personal protection, but they may also increase the potential for dengue transmission. Supporters may inadvertently introduce into Brazil new dengue genotypes to which local immunity is low, and the assembly of large non-immune, and hence susceptible, populations could fuel transmission in the event of an outbreak.

> ## FIFA, THE BRAZILIAN AUTHORITIES AND THE WORLD CUP SPONSORS MUST USE THEIR INFLUENCE AND EXPERIENCE TO COMMUNICATE THE RISK.

Seasonal averages, by definition, are not always an accurate guide to risk. It will also be prudent to monitor dengue outbreaks in the time leading up to and during the World Cup. This can be done using online resources such as DengueMap (www.healthmap.org/dengue) and Dengue Trends (www.google.org/denguetrends). DengueMap is a collaboration between the US Centers for Disease Control in Atlanta, Georgia, and HealthMap, an automated, web-based monitoring and reporting system. Founded by a team at the Boston Children's Hospital in Massachusetts, HealthMap collects, collates and maps formal and informal reports of dengue outbreaks to provide a free guide to them. Google's Dengue Trends reports on the volume of Google searches for dengue in a given location, a potential proxy for increased risk.

The World Cup is an opportunity to evaluate the uptake of these new public-health information systems and their utility both to individuals and the authorities. Crucially, if they can provide timely feedback on the effectiveness of preventive measures for the authorities on the ground, that could prompt yet further responses.

I don't want to dissuade anyone from going to the World Cup, nor to single out Brazil, which is just one of more than 100 countries worldwide battling dengue. My aim is to inform unwary spectators about the risk and how they can protect themselves, and how the risk could be mitigated by on-the-ground control measures.

PS Come on England! ∎

↻ **NATURE.COM**
Discuss this article
online at:
go.nature.com/bjrzfv

Simon Hay *is a Wellcome Trust fellow at the University of Oxford, UK. e-mail: simon.hay@zoo.ox.ac.uk*

PHARMACOLOGY

Painkiller kills the bad effects of pot

Marijuana's undesirable effects on the brain can be overcome by using painkillers similar to ibuprofen, at least in mice.

Chu Chen and his colleagues at Louisiana State University Health Sciences Center in New Orleans treated mice with THC, marijuana's main active ingredient. They found that THC impaired the animals' memory and the efficiency of their neuronal signalling, probably by stimulating the enzyme COX-2.

The authors reversed these negative effects — and were able to maintain marijuana's benefits, such as reducing neurodegeneration — when they also treated the mice with a drug, similar to ibuprofen, that inhibits COX-2.

The authors suggest that the benefits of medical marijuana could be enhanced with the use of such inhibitors.
Cell 155, 1154–1165 (2013)

PALAEOECOLOGY

Dung reveals goats' last days

Climate change, rather than human actions, probably drove the Balearic mountain goat (*Myotragus balearicus*) extinct.

This small goat, unique to Spain's Balearic Islands in the western Mediterranean, disappeared soon after

CLIMATE SCIENCES

Crusty alga uncovers sea-ice loss

Like tree rings, layers of growth in a long-lived Arctic alga may preserve a temperature record of past climate. Specimens from the Canadian Arctic indicate that sea-ice cover has shrunk drastically in the past 150 years — to the lowest levels in the 646 years of the algal record.

Satellite records of the Arctic's shrinking sea-ice cover date back only to the late 1970s. Jochen Halfar of the University of Toronto at Mississauga, Canada, and his colleagues have found a new palaeoclimate proxy in the coralline

marine alga *Clathromorphum compactum* (pictured). It can live for hundreds of years and builds a fresh layer of crust each year.

The thickness of each layer, and the ratio of magnesium to calcium within it, are linked to water temperature and the amount of sunlight the organism receives. The discovery suggests a new way to calculate how much polar sea ice existed hundreds of years ago.
Proc. Natl Acad. Sci. USA http://doi.org/p6g (2013)

humans arrived on the islands, about 5,000 years ago. Some researchers have proposed that disease or hunting by humans killed off the goats.

Frido Welker and Barbara Gravendeel of the Naturalis Biodiversity Center in Leiden, the Netherlands, and their colleagues analysed plant DNA found in the goats' fossilized faeces (pictured). The results suggest that the goats were dependent on *Buxus balearica*, a local species of shrub. Further analysis indicated that the shrub's abundance on the islands

declined sharply 4,000–5,000 years ago because of a drier climate. This is likely to have contributed greatly to the goats' extinction.
Quat. Res. http://doi.org/p6b (2013)

NEUROSCIENCE

Satiety signal from the mouth

A human hormone might be a potent treatment for obesity, but only if it is taken orally.

The peptide hormone PYY is made primarily by cells in

the gut as a satiety signal to the brain. When it is injected into humans, however, it causes nausea and ruins the taste of food. Sergei Zolotukhin at the University of Florida in Gainesville and his colleagues sprayed PYY into the mouths of mice and found that although the animals stopped eating, as expected, they did not become nauseous.

PYY in saliva seems to use a different signalling pathway from gut PYY to tell the brain when it is time to stop eating. Targeting molecules in this pathway with oral PYY or

other compounds could reduce overeating without inducing nausea.
J. Neurosci. 33, 18368–18380 (2013)

VIROLOGY

A more predictable flu

Influenza viruses may have fewer routes for escaping vaccines than previously thought.

Flu vaccines target a viral protein called haemagglutinin, which mutates frequently, rendering vaccines ineffective. Derek Smith of the University of Cambridge, UK, and Ron Fouchier of Erasmus Medical Center in Rotterdam, the Netherlands, and their colleagues studied how the haemagglutinin protein has mutated to evade vaccines — a process called antigenic drift — over a 35-year period from 1968 to 2003.

They found that seven of the ten antigenic drift events in the past three decades were caused by a change in just one amino acid in the protein. These changes occurred at only seven places in the protein, all of which cluster near a region that binds to host cells. The results could one day lead to more-effective flu vaccines.
Science 342, 976–979 (2013)

ECOLOGY

Mother frogs arm their tadpoles

Some animals make chemical defences against predators; others obtain them from their food. Researchers have now found the first example of parents chemically arming their young after birth.

Ralph Saporito of John Carroll University in University Heights, Ohio, and his colleagues analysed specimens of the strawberry poison frog *Oophaga pumilio* (**pictured**) from all stages of the life cycle. Newly hatched tadpoles had no defensive alkaloids, but after their mothers began producing unfertilized 'nutritive eggs' for

them to eat, tadpole alkaloid concentrations rose. Adult frogs obtain the alkaloids from ants and mites in their diet.

Hand-reared *O. pumilio* that were fed nutritive eggs from another frog species lacking chemical defences remained alkaloid-free.
Ecology http://doi.org/p59 (2013)

PHYSIOLOGY

Smells maintain blood cells

Fruitfly larvae need to sense odours to maintain a pool of the cells that give rise to blood cells.

A team led by Utpal Banerjee at the University of California, Los Angeles, studied mutants of *Drosophila melanogaster* to identify molecular signals connecting odour sensing to blood progenitor cells. They found that smells prevent the cells from specializing, or differentiating, before they are required.

When the team activated olfactory neurons in the fly's brain, the neurons secreted a chemical called GABA into the blood, triggering blood progenitors to let in calcium ions. Calcium maintains the cells as undifferentiated progenitors. Larvae raised in environments with few odours had low levels of GABA, and their blood progenitor cells differentiated earlier.

Whether similar links exist between sensory perception and progenitor cells in more complex organisms is not clear.
Cell 155, 1141–1153 (2013)

PHYSICS

Pinch of salt makes for bumpy icicles

⭐ **HIGHLY READ** on iopscience. iop.org 20 Oct–19 Nov

Impurities in water are behind the ripples seen around an icicle's circumference.

Ripples or ribs form naturally in icicles, an effect that previous theories attributed to surface tension in the thin film of water that flows over the ice. Antony Szu-Han Chen and Stephen Morris at the University of Toronto, Canada, analysed 67 icicles grown under a broad range of conditions in the laboratory. They found that whereas icicles made from pure water were ripple-free, even small amounts of salt dissolved in the water — less than is found in most tap water — caused ripples to emerge. The ribs also grew faster in saltier water.

Existing theories do not account for the effects of impurities in ripple formation, leaving salt's role in the process a mystery.
New J. Phys. 15, 103012 (2013)

GLACIOLOGY

Anatomy of an ice shelf's demise

The sudden drainage of thousands of small lakes on the surface of Antarctic glaciers seems to have triggered the spectacular collapse of the Larsen B ice shelf in March 2012.

Some 3,000 small ponds of liquid water had emerged over the course of a decade on top of glaciers surrounding the ice shelf on the Antarctic Peninsula. These ponds disappeared in striking synchronicity a few days before the shelf's collapse.

When recreating the events in a computer simulation, Alison Banwell of the University of Chicago in Illinois and her colleagues found that the initial drainage of a single lake would have produced fractures in the ice that were capable of sucking dry neighbouring lakes, kicking off a catastrophic chain reaction.

The spread of fractures across the ice shelf may have ultimately

caused its sudden demise, the authors suggest.
Geophys. Res. Lett. http://doi.org/p6c (2013)

ORGANIC CHEMISTRY

Fast and easy fluorine fix

Many drug compounds and agrochemicals are fluorinated, but adding fluorine atoms to organic molecules can be dangerous and expensive. Patrick Fier and John Hartwig at the University of California, Berkeley, report a way to fluorinate one class of molecules at room temperature and without the need for harsh reagents.

They showed that silver(II) fluoride can swap a hydrogen atom for a fluorine atom on molecules containing nitrogen as part of a ring of carbon atoms. The reaction replaces only the hydrogen attached to carbons next to the nitrogen in the ring. It occurs quickly and uses only commercially available reagents.
Science 342, 956–960 (2013)

↻ **NATURE.COM**
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

UN climate talks

International climate negotiators signed a landmark agreement on forest conservation at the United Nations climate summit in Warsaw on 23 November. The forest framework will enable carbon payments to countries that can document a reduction in deforestation. Negotiators also established a 'loss and damage' mechanism that will help poor countries to pay for the impacts of global warming. Progress on a new climate treaty scheduled for signing in Paris in 2015 was minimal, although countries agreed to submit commitments by early that year. See go.nature.com/pfswg2 for more.

Stem-cell laws

Japan passed two regenerative-medicine laws on 20 November. A law intended to hasten approval of stem-cell treatments rules that candidate therapies no longer need to pass rigorous phase III clinical trials to prove efficacy — a move that critics say could flood the market with ineffective treatments (see *Nature Med.* **19,** 510; 2013). Another law requires physicians to report the clinical use of unapproved stem-cell therapies to the Japanese health ministry. The new regulations will be implemented within a year.

Research waste

Valuable, large-scale scientific instruments are standing idle in the United Kingdom because of poor financial planning, according to a parliamentary report released on 21 November. The House of Lords Science and Technology Committee cited a "damaging disconnect" between the funding provided to build new research facilities,

ASAHI SHIMBUN/GETTY

Eruption raises Japanese island

Lava from an underwater volcano breached the surface of the Pacific Ocean last week, forming an island (pictured, right) some 1,000 kilometres south of Tokyo. The islet is about 200 metres in diameter and lies off the coast of Nishinoshima, an uninhabited island in the Ogasawara chain, or Bonin Islands. They and the rest of the Japanese archipelago form part of the seismically active 'Ring of Fire'. In the past, other small islands have appeared near Japan and then eroded; officials are waiting to see if the new land mass persists before naming it. In September, an earthquake in Pakistan created an island off the country's Gwadar coast, measuring about 50 metres by 20 metres (see *Nature* **502,** 10–11; 2013).

and provision of the resources needed to operate them. For example, the government spent nearly £40 million (US$65 million) on a high-performance-computing centre, without budgeting for the electricity required to run the computers. See go.nature.com/6bw4bh for more.

Iran nuclear deal

Iran agreed on 24 November in Geneva, Switzerland, to curtail its nuclear programme temporarily in exchange for partial relief from international sanctions. In a historic pact with the United States, the United Kingdom, Germany, France, Russia and China, the country has committed for six months to keep its uranium enrichment to below 5% purity (well short of the threshold for weapons-making), to neutralize stockpiles of more highly enriched uranium, and to allow frequent inspections by the Vienna-based International Atomic Energy Agency.

Horizon 2020

The European Parliament on 21 November formally approved Europe's Horizon 2020 programme — a roughly €80-billion (US$108-billion) scheme to fund research and innovation. The parliament had initially hoped to budget €100 billion for the initiative, but its proposal was rejected in February by the heads of European Union (EU) member states. Horizon 2020, which runs from 2014 to 2020, introduces simplified funding rules to reduce red tape in EU-funded research (see *Nature* http://doi.org/p7n; 2013).

Science ship boost

A key research vessel for international ocean drilling is getting a new lease of life. The US National Science Foundation (NSF) received authorization on 21 November to provide the *JOIDES Resolution* with as much as US$250 million in funding over the next five years — an amount to be supplemented

with some $87.5 million from international partners. Amid tightening budgets, the NSF had previously considered slashing money for the vessel (see *Nature* **501**, 469–470; 2013). Final funding levels must still be decided by Congress. See go.nature. com/5sbcqd for more.

BUSINESS

DNA diagnosis

The US Food and Drug Administration has approved the first devices that use next-generation DNA-sequencing for diagnosis. The decision, made on 19 November, includes three kits to be used with the MiSeqDx sequencer made by Illumina in San Diego, California. Two of them pinpoint genetic mutations linked to cystic fibrosis, and the third would allow clinical labs to develop their own tests for other disorders. See go.nature. com/auzrmg for more.

Arrested testing

Personal-genetics firm 23andMe in Mountain View, California, must stop marketing its genetic testing service, says the US Food and Drug Administration (FDA). In a warning letter on 22 November, the agency said that the saliva-testing kit has yet to be approved or cleared as a medical device. The company filed for FDA clearance

of several tests for various diseases in 2012, but the agency said that 23andMe has not since provided the data to validate those tests, nor others covered by its kit, and has not responded to communications since May. See go.nature.com/uozdie for more.

PEOPLE

Philanthropist dies

Philanthropist and science advocate Fred Kavli died on 21 November, aged 86. He established the Kavli Foundation in Oxnard, California, in 2000 to support research and promote public understanding of science. The foundation has established research institutes internationally, and awards the biennial Kavli prizes — three US$1-million awards in the areas of astrophysics, nanoscience and neuroscience. The Kavli Foundation also had a central role in jump-starting the US BRAIN Initiative (see *Nature* **503**, 26–28; 2013).

Nobel laureate dies

Two-time Nobel prizewinner Frederick Sanger (**pictured**) died on 19 November, aged 95. Sanger won the 1958 Nobel Prize in Chemistry for developing a method that let him determine the complete amino-acid sequence of insulin — a method that has

since been widely applied to other proteins. In 1980, he won the chemistry prize again, for discovering a technique for sequencing DNA. An adapted version of Sanger sequencing was later used to decode the human genome for the first time. See go.nature.com/lfc7yt for more.

AWARDS

Mentoring prize

This year's *Nature* Awards for Mentoring in Science have been given to three scientists in Italy: neurobiologist Michela Matteoli of the University of Milan, physicist Giorgio Parisi of the Sapienza University of Rome, and chemist Vincenzo Balzani of the University of Bologna. Italy's President Giorgio Napolitano presented the prizes on 25 November at the Quirinal Palace in Rome. Each year, the awards honour outstanding scientific mentors in a different country or region. See pages 438 and 559 for more.

COMING UP

3–5 DECEMBER
Experts convened by the World Health Organization will select a shortlist from 22 proposals for using new ways to pay for health technologies in poor countries; winners chosen next year will run as demonstration projects.
go.nature.com/4uovpx

EVENTS

Solar sensor launch

A US satellite designed to measure the Sun's energy output launched on 19 November from NASA's Wallops Flight Facility on Wallops Island, Virginia. Data collected by the Total Solar Irradiance Calibration Transfer Experiment (TCTE) will help researchers to understand the Sun's impact on Earth's climate. The TCTE, funded by the US National Oceanic and Atmospheric Administration, is intended to replace the ageing Solar Radiation and Climate Experiment, which has been in orbit for more than ten years — twice its intended lifespan (see *Nature* **469**, 457–458; 2011).

Swarm takes flight

The European Space Agency has launched a trio of satellites designed to survey Earth's magnetic field. The €220-million (US$296-million) Swarm mission lifted off on 22 November from the Plesetsk Cosmodrome in northern Russia. The mission, slated to last for four years, will measure temporal and spatial variations in the planet's magnetic field in unprecedented detail. See go.nature.com/zi2fkk and *Nature* http://doi.org/p7p (2013) for more.

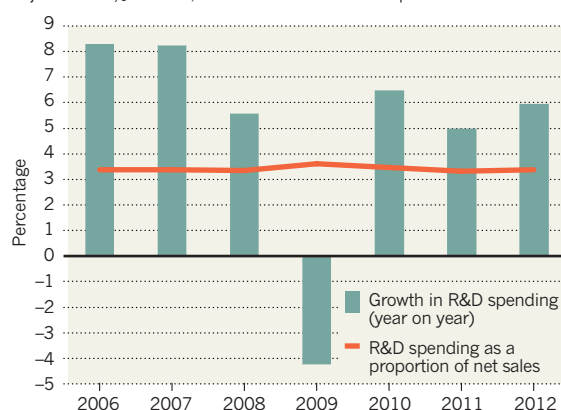↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

The 2,000 companies that spend the most on research and development increased their investment by 6.2% last year, according to the European Commission's 2013 Industrial R&D Investment Scoreboard. This is broadly in line with changes in net sales. German car-maker Volkswagen, headquartered in Wolfsburg, was the world's biggest research investor with €9.5 billion (US$12.9 billion) spent in 2012; electronics firm Samsung in Seoul ranked second with €8.3 billion.

SLOW RECOVERY FOR CORPORATE RESEARCH

Global spending by firms on research and development (R&D) remains at just over 3% of sales, and has not recovered to pre-recession levels.

Legend:
- Growth in R&D spending (year on year)
- R&D spending as a proportion of net sales

y-axis: Percentage (−5 to 9)
x-axis: 2006, 2007, 2008, 2009, 2010, 2011, 2012

NEWS IN FOCUS
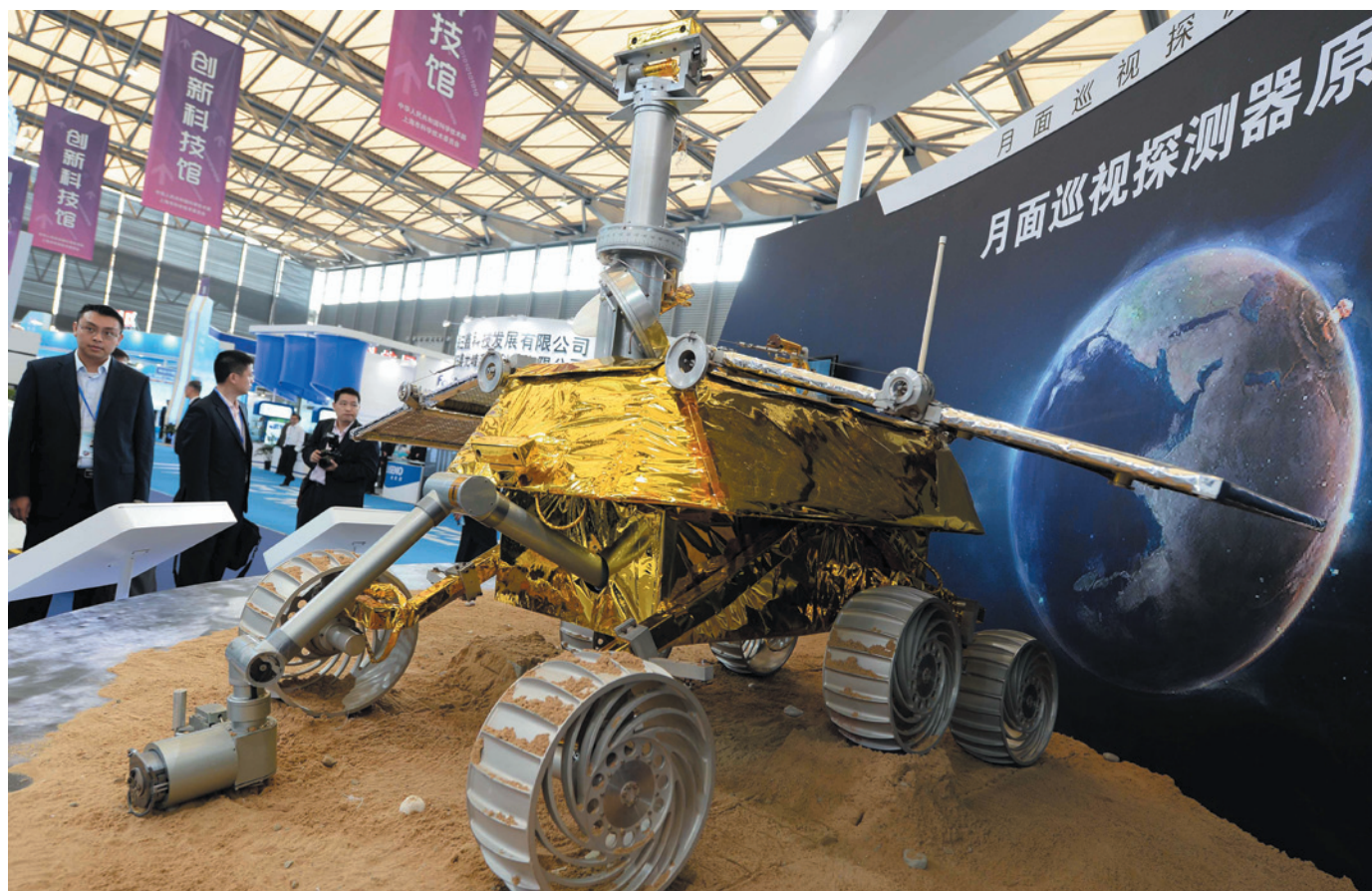
Preparing for lift-off: China displays a model of the Yutu rover that it plans to send to explore the Moon in December.

SPACE

China aims for the Moon

Planned launch of lunar rover follows a string of triumphs for the country's space programme.

BY ALEXANDRA WITZE

Next month, a Chinese spacecraft called Chang'e-3 is scheduled to use braking rockets to lower itself gently onto the plains of Sinus Iridum, a broad swathe of lava flows on the near side of the Moon. The probe will then roll out a six-wheeled rover — the first machinery to explore the Moon's surface since 1976, when the Soviet Luna 24 mission scooped up a handful of soil and flew it back to Earth.

The landing would be the latest step in China's methodical and almost flawless space programme. The country has achieved a string of triumphs in crewed space flight over the past decade, including putting humans into orbit and docking two craft in space. China lost its first and only Mars probe soon after launch in 2011, but both of its lunar orbiters flew successfully.

If Chang'e-3 lands safely on the Moon, China will join the Soviet Union and the United States as the only nations to have successfully landed exploratory spacecraft there. "You cannot call the Chinese a rising or emerging space power any more," says Bernard Foing, a lunar scientist at the European Space Agency in Noordwijk, the Netherlands. "They have shown they are very advanced."

The roots of China's lunar programme trace back to the early 1990s, when money began to flow into work on crewed space flights and space scientists pushed for a parallel programme in lunar exploration. The result was a schedule of missions named after Chang'e, a luminescent Moon goddess.

Chang'e-1, an orbiter launched in 2007 by the Beijing-based China National Space Administration, mapped the entire Moon before it was deliberately crashed into the lunar surface in 2009. Chang'e-2, launched in 2010, made higher-resolution maps before moving on to fly past the asteroid Toutatis, which it did last December.
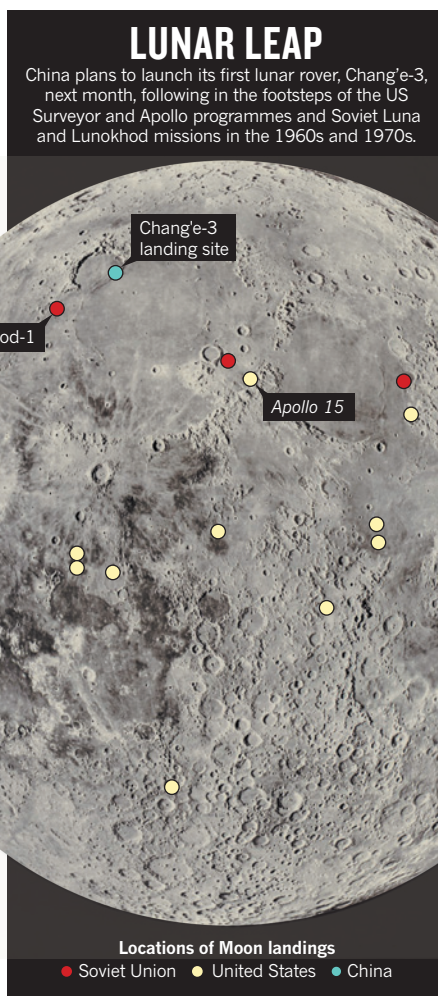
Chang'e-3 is slated as the first step of China's second phase of exploration. The probe is expected to launch from the Xichang launch centre in Sichuan province in December.

If the mission launches on 1 December, Chang'e-3 could enter into lunar orbit on 6 December, says Foing. The probe could then land in Sinus Iridum in the Moon's mid-latitudes on 16 December. Also known as the Bay of Rainbows, this location is close to where the Soviet Lunokhod-1 mission trundled in 1970–71, and on the opposite side of the great Mare Imbrium basin from where the US *Apollo 15* mission landed (see 'Lunar leap').

"There's nothing particularly interesting about the spot, but it is a place we haven't been to before," says Paul Spudis, a Moon researcher at the Lunar and Planetary Institute in Houston, Texas. Sinus Iridum is also a fairly safe place to land, with flat plains and relatively few boulders.

If Chang'e-3 makes it to the surface, the lander portion will remain in one spot, kept warm during the frigid lunar nights by a radioactive heat source. It will survey Earth, the Milky Way and the rest of the sky with the first near-ultraviolet telescope ever deployed on the Moon, which will help astronomers to observe the birth and death of stars.

The solar-powered, 100-kilogram rover — named Yutu, or 'jade rabbit' — is expected to explore the vicinity. Panoramic and other cameras will photograph the surroundings, and an α-particle X-ray spectrometer on a robotic arm will probe the soil's chemical composition. Ground-penetrating radar will also scan the Moon's subsurface to depths of 100 metres or more to study soil and rock structures, says Wenzhe Fa, a remote-sensing specialist at

LUNAR LEAP

China plans to launch its first lunar rover, Chang'e-3, next month, following in the footsteps of the US Surveyor and Apollo programmes and Soviet Luna and Lunokhod missions in the 1960s and 1970s.

Chang'e-3 landing site

Lunokhod-1

Apollo 15

Locations of Moon landings
● Soviet Union ● United States ● China

Peking University in Beijing.

Depending on what happens with Chang'e-3, the National Space Administration may launch an almost identical rover and lander pair — Chang'e-4 — to another spot on the lunar surface. Beyond that, the third and final phase of China's lunar-exploration programme calls for a robotic mission to bring back samples of lunar material, probably in 2017–18.

Space analysts expect that the lunar and crewed objectives of China's space-flight programme will merge, with Chinese astronauts (known as taikonauts) aiming to walk on the Moon some time in the 2020s. China's plans are notable for their long-term outlook — not so easy to implement in a democracy — and for proceeding incrementally, says Joan

Johnson-Freese, an analyst at the US Naval War College in Newport, Rhode Island. "They have a long laid-out programme of very careful steps, but they are taking bigger steps with each flight," she says.

For the moment, those steps surpass what any other country is doing. South Korea recently announced that it would send an uncrewed lander to the Moon in 2020 (see *Nature* http://doi.org/p6h; 2013), but those plans — like others — remain just that, plans. Russia is considering the development of a series of orbiters and landers, and Japan has discussed a lander and rover, but schedules and budgets remain unclear. The United States has nothing lined up to follow its two current Moon orbiters, the Lunar Reconnaissance Orbiter and the Lunar Atmosphere and Dust Environment Explorer (LADEE) that is studying dust in the atmosphere.

In fact, LADEE faces a challenge when Chang'e-3 arrives at the Moon. The Chinese orbiter is expected to release large amounts of exhaust gases when it enters lunar orbit, which LADEE will have to sort through to separate from naturally occurring dust. But NASA scientists cannot work on this task directly with those overseeing Chang'e-3 in Beijing because legislation pushed through by US Representative Frank Wolf (Republican, Virginia) forbids bilateral collaboration between the US agency and Chinese scientists.
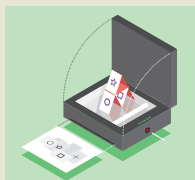
However, others are moving forward with international agreements. The Paris-based European Space Agency will hold a meeting in February in Chengdu, China, to explore possible future joint missions involving the Chinese National Space Science Center in Beijing, which oversees the country's space-science research. The two have already collaborated on one satellite project: the Cluster/Double Star mission in 2003–07 to study Earth's magnetosphere. The National Space Science Center is also gearing up to launch China's first X-ray satellite in 2015, says its director, Wu Ji.

"They're running their own race," says Johnson-Freese. "They're not looking behind them." ∎

Data from the Large Hadron Collider, such as this decay of a Higgs boson, could be made publicly available.

PHYSICS

LHC plans for open data future

Researchers share results to keep them accessible.

BY ELIZABETH GIBNEY

When the Large Hadron Collider (LHC) is humming along, the data come in a deluge. The four experimental detectors at the facility, based at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, collect some 25 petabytes of information each year.

Storing the data is not a problem: hard drives are cheap and getting cheaper. The challenge is preserving knowledge that is less commonly stored — the software, algorithms and reference plots specific to each experiment. These often degrade or disappear with time, says Cristinel Diaconu of the Marseilles Centre for Particle Physics in France, who is chair of the international Data Preservation in Long Term Analysis in High Energy Physics (DPHEP) study group. He worries that if the data continue to be stored in their current state, physicists trying to decipher them in 10 years' time will be unable to reconstruct the discovery of the Higgs boson. "When the LHC programme comes to an end, it will probably be the last data at this frontier for many years," he says. "We can't afford to lose it."

The DPHEP is therefore trying to push data-preservation efforts from mere storage to a system of open sharing. The thinking goes that data and the knowledge needed to interpret them are more likely to survive in the long term if many people outside an experiment are constantly trying to make sense of them.

Kati Lassila-Perini, a physicist at the Compact Muon Solenoid (CMS), one of the four experiments at the LHC, has a radical idea for this sort of sharing: giving data away to school pupils. Next year, a pilot scheme she leads will release 2010 CMS data, which the IT Center for Science in Espoo, Finland, will reformat and store. The centre will then share the data with pupils, who will recreate plots of particle decays using analysis tools adapted for the public. The CMS plans to make more data publicly available a few years after collection, and Lassila-Perini hopes that other data centres will adopt such schemes. "We are guaranteeing that the data we are not looking at any more remain accessible," she says.

The intent is not just to keep data for posterity. Old data can be mined to test new theories and provide crucial references for new experiments, says Diaconu. Before the Higgs boson was discovered in 2012, for example, the Large Electron–Positron collider — the LHC's predecessor at CERN — came back into the spotlight as physicists scoured its 1990s-era data, looking for an exotic type of Higgs that had not been theorized at the time the data had been gathered. In this way, the goals of keeping data alive and open are "enlightened self-interest", says Michael Hildreth, a physicist at

the University of Notre Dame in Indiana and leader of the US-funded Data and Software Preservation for Open Science (DASPOS) effort, which has similar goals to the DPHEP.

DASPOS is building a template for preserving data — a checklist of items that should be stored, and how to do it. Next year, in a 'curation challenge', DASPOS will task physicists with recreating results from other experiments using only the information collected with this template. One test will almost certainly use LHC data — challenging, for example, CMS physicists to recreate results from the rival ATLAS experiment. Another test could come from a different field, such as astrophysics. If successful, the model could form a generic and simplified architecture for preserving data, says Hildreth.

Part of the challenge is coping with ever-changing algorithms, operating systems and data-analysis hardware. At the German Electron Synchrotron (DESY) in Hamburg, computing coordinator David South is leading a project that is already attempting to protect data in this way. His team has devised a system that will automatically comb through data and software from experiments on DESY's Hadron–Electron Ring Accelerator and test them for compatibility when hardware or operating systems change.

This plan to migrate data repeatedly onto new platforms stands in contrast to an approach at the BaBar experiment at the SLAC National Accelerator Laboratory in Menlo Park, California. There, versions of data and the operating systems needed to analyse them have been frozen in storage centres, where they are supposed to be accessible until at least 2018. South says that DESY's approach is more reliable. Although DESY's system needs monitoring — any incompatibilities must be fixed through human intervention — the goal is to deal with problems as they arise, rather than tackle them years later, when they may have compounded.

> *"When the LHC programme comes to an end, it will probably be the last data at this frontier for many years. We can't afford to lose it."*

DESY scientists would know about that. In the 1990s, physicists wanted to take another look at data from a DESY collider that ran from 1979 to 1986, to further investigate the strong interaction that binds quarks together. They managed to measure it with increased precision, but Diaconu says that it took two years to reconstruct the data, which had not been maintained.

The data preservationists are quick to point out the expense associated with reconstruction efforts. Of course, preservation also costs money, but it is well worth it, says DPHEP project manager Jamie Shiers. He puts the bill for implementing good data-preservation at the LHC at around 1% of operating costs — just a few million dollars per year. "I think it's justified," he says. ∎

DRUG DISCOVERY

Project ranks billions of drug interactions

Drugable.com predicts mechanisms through computation.

BY SARA REARDON

For decades, drug development was mostly a game of trial and error, with brute-force candidate screens throwing up millions more duds than winners. Researchers are now using computers to get a head start. By analysing the chemical structure of a drug, they can see if it is likely to bind to, or 'dock' with, a biological target such as a protein. Such algorithms are particularly useful for finding potentially toxic side effects that may come from unintended dockings to structurally similar, but untargeted, proteins.

Last week, researchers presented a computational effort that assesses billions of potential dockings on the basis of drug and protein information held in public databases. "It's the largest computational docking ever done by mankind," says Timothy Cardozo, a pharmacologist at New York University's Langone Medical Center, who presented the project on 19 November at the US National Institutes of Health's High Risk–High Reward Symposium in Bethesda, Maryland. The result, a website called Drugable (drugable.com) that is backed by the US National Library of Medicine (NLM), is still in testing, but it will eventually be available for free, allowing researchers to predict how and where a compound might work in the body, purely on the basis of chemical structure (see 'Mining for drugs').

Cardozo acknowledges that the computations are just an initial step in drug discovery. After predicting whether a protein can bind to a compound, drug developers must test the drug's action on the same protein in a cell to see what actually happens to the protein's function, as well as how much of the drug is needed and under what conditions. Then come animal trials and, if researchers are lucky, human trials. But these extra data are often proprietary and held by pharmaceutical companies, says Brian Shoichet, a computational biologist at the University of California, San Francisco. Some public databases such as PubChem, maintained by the NLM, hold the results of automated tests of drugs on proteins in yeast cells, but they contain inaccuracies and false positives, he says.

Still, scientists have already shown that the computational approach can provide some short cuts. In 2012, Shoichet and researchers at the Novartis Institutes for BioMedical Research in Cambridge, Massachusetts, developed an algorithm that predicts side effects on the basis of similarities between drugs' chemical structures. When the researchers tested the program on 656 approved drugs and 73 biological targets, they found that it predicted hundreds of previously unknown interactions — and that these side effects turned out to be real about half of the time (E. Lounkine *et al. Nature* **486,** 361–367; 2012). For known drugs, Shoichet says, this type of computation provides a ▶

MINING FOR DRUGS
Researchers are using Google supercomputers to examine billions of drug–protein interactions.
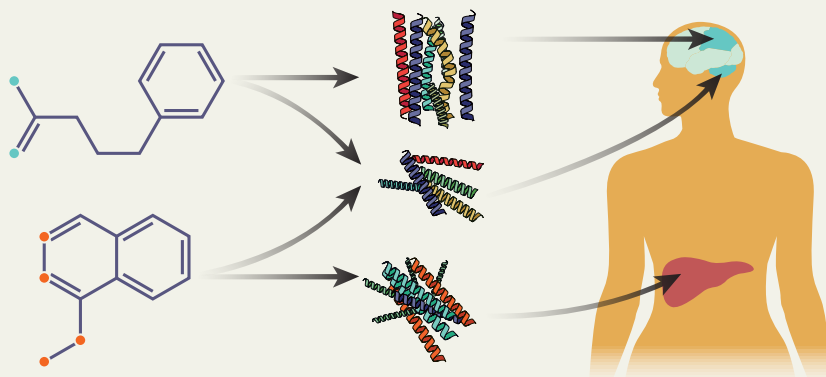
1 Researchers obtained publicly available profiles of 600,000 drugs and chemical compounds.

2 They evaluated each molecule's abilities to bind to some 7,000 chemical 'pockets' on 570 human proteins.

3 They looked at the bodily expression of those proteins to predict where the drugs' effects might be seen.

▶ quick way to identify interactions that should be investigated further.

Predicting how untested compounds will interact with proteins in the body, as Drugable attempts to do, is more challenging. In setting up the website, Cardozo's group selected about 600,000 molecules from PubChem and the European Bioinformatics Institute's ChEMBL, which together catalogue millions of publicly available compounds. The group evaluated how strongly these molecules would bind to 7,000 structural 'pockets' on human proteins also described in the databases. Computing giant Google awarded the researchers the equivalent of more than 100 million hours of processor time on its supercomputers for the mammoth effort.

The team came up with ranked docking scores describing some 4 billion potential drug–protein interactions. Then the group cross-referenced the target proteins with those in the NLM's Gene Expression Omnibus database, which shows where in the body different genes that code for proteins are expressed. This allowed them to predict where the drug might act, says Cardozo: if Drugable finds an interaction for a protein that is highly expressed in a certain tissue, chances are good that the effect would manifest itself in that tissue.

Pharmaceutical companies have been doing similar computational predictions for years, says Jeremy Jenkins, a researcher at the Novartis Institutes. But he says that Novartis, which has a library of 1.5 million public and proprietary compounds, has never attempted to analyse as many proteins and drugs at once as Drugable has done.

Cardozo hopes that Drugable will be particularly helpful in evaluating psychiatric drugs, which often act in ways that are difficult to measure. As a demonstration, Cardozo's group applied Drugable's algorithm to clozapine and chlorpromazine, two drugs often prescribed to treat schizophrenia.

As expected, Drugable showed that the two drugs bind most strongly to receptors for the neurotransmitters serotonin and dopamine, which are expressed in the parts of the brain involved in higher information processing. But it found that clozapine, which also stabilizes mood disorders such as depression, binds strongly to a particular dopamine receptor called DRD4, which is expressed in the brain's pineal gland — a known mood regulator.

The group also found that clozapine binds to a receptor in the part of the brain that regulates saliva production; excessive salivation is a known side effect of clozapine. Although the biochemical explanations for mood regulation and salivation have been proposed before, Cardozo says that Drugable can be used to reveal the most plausible mechanisms. ∎

The red turpentine beetle has wiped out more than 10 million pine trees in China in the past 15 years.

ECOLOGY

China battles army of invaders

Raft of control measures slows the march of alien species.

BY JANE QIU IN QINGDAO

When China kicked open the doors to international trade in the late 1970s, not everything that came in was welcome. Along with Western goods and new technologies, alien organisms infiltrated the country. A comprehensive survey now reveals that almost 550 non-native species, from viruses to fish and mammals, have become invasive in the country (see 'Space invaders'). They are costing an estimated US$15 billion in losses each year, with damage to crops and forests a particular problem.

"As the volume of international trade has grown exponentially, so has the number of alien species," said Li Bo, director of the Office for Management of Alien Species in the Ministry of Agriculture, Beijing, at the second International Congress on Biological Invasions in Qingdao last month.

Since 2000, China has tightened its regulations on importing plant materials and has enforced strict quarantine requirements. It has also spent more than $1 billion establishing databases of invasive species, monitoring their spread, researching invasive mechanisms and ecological impact, and developing control technologies. This has led to an "explosion of research", says Wan Fanghao, an ecologist at the Chinese Academy of Agricultural Sciences' Institute of Plant Protection in Beijing.

Wan is currently finalizing a $10-million, decade-long project funded by the Ministry of Science and Technology to study invasive species in agriculture and forestry. At the Qingdao meeting, where some of the results were presented, scientists showed how a better understanding of what makes an alien species invasive could aid in the development of effective controls.

A case in point is the whitefly *Bemisia tabaci*, an insect that feeds on plant vascular tissue called phloem. It causes damage both directly through feeding and indirectly through the transmission of plant viruses, and has wreaked havoc on vegetable and cotton production in all of China's provinces except Tibet. In a major outbreak in 2009, "a quarter of vegetable farms nationwide, about 200,000 hectares, were plagued, which reduced the yields by 50% to 80%", says Liu Shusheng, an entomologist at Zhejiang University in Hangzhou.

Researchers have now managed to halt the whitefly's march. Strategies such as planting crop varieties that are resistant to the pest, separating individual seedlings to minimize pest spread, applying low levels of pesticides and implementing biological control with natural enemies means that "there haven't been major

outbreaks since 2009", says Wan.

Another invader that has been brought under control is the red turpentine beetle (*Dendroctonus valens*). In North America, the beetle mainly attacks dead or ailing trees. But the beetles, which were introduced to China in the 1980s, have wiped out more than 10 million pine trees in northern provinces since 1999.

A study led by Sun Jianghua, an entomologist at the Chinese Academy of Sciences' Institute of Zoology in Beijing, found that the interaction between the beetles and their symbiotic fungus *Leptographium procerum* is key to their 'personality change' in China (J. Sun *et al. Annu. Rev. Entomol.* **58,** 293–311; 2013). Since its arrival, "the fungus has mutated into novel genotypes", says Sun. One of these induces trees to release large amounts of the compound 3-carene — a strong attractant to the beetles — that is not released in response to the North American fungal variant.

The finding has led to a series of successful projects to trap beetles using 3-carene. The approach, says Sun, is part of an integrated pest-management programme, launched in 2007, that also includes the use of other chemical attractants and pesticides, and efforts to replace single-species forests with a mix of plants.

As a result, the spread of the red turpentine

SPACE INVADERS

China currently has 544 invasive species, a fourfold increase on the number in 1900.
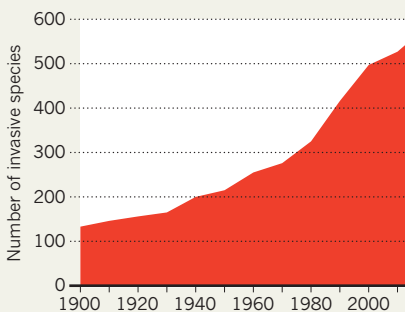
beetle is mostly under control, says Sun. Fewer than 1 in 1,000 trees are now infected, compared with the staggering 3 in 10 that were affected in Shanxi province in 2001, during one of the worst outbreaks.

Sun's findings raise the possibility of a potential 'reinvasion' of the United States by the red turpentine beetle and its Chinese fungal variant, says Daniel Simberloff, an ecologist at the University of Tennessee in Knoxville. "The policy implications are huge," he adds.

What is happening in China matters to the rest of the world, says Helen Roy, an ecological entomologist at the Centre for Ecology and Hydrology in Wallingford, UK. Biological invasions are "two-way traffic", she says. Most of the forest pest species in North America originally came from China and some of its exports have wreaked havoc in Europe.

When dealing with invasive species, international collaboration is extremely important, says Roy. She has been studying the invasion of the harlequin ladybird (*Harmonia axyridis*) in Europe and, by working with Chinese researchers, is now trying to understand the insect's behaviour and natural enemies in the hope of developing effective control measures.

But administrative issues in China sometimes hamper scientists' efforts, says Wan. For instance, many alien species enter China by piggybacking on imports of rubbish from developed countries (waste disposal is big business in China). But it is unclear which ministry is in charge of inspection and monitoring of the cargo. Moreover, tackling invasive species often involves multiple ministries. "There needs to be better coordination and more data sharing between them," says Sun.

In any case, the problem of invasive species will not go away, says Wan. "With climate warming, increasing international trade and rapid urbanization, the problem of biological invasions will only get worse," he says. "We need to keep a close eye on potential troublemakers." ∎

CONSERVATION

Nations fight back on ivory

Politicians take action on poaching in Africa as tusk seizures approach record numbers.

BY DANIEL CRESSEY

It has been a bad year for Africa's elephants. Thousands have been killed as poachers rush to cash in on soaring ivory prices, which have reached hundreds of dollars per kilogram. The cyanide poisoning of up to 300 animals at watering holes in a game park in Zimbabwe last month served as a particularly unpleasant reminder of the lengths to which poachers are willing to go.

Official numbers for elephant killings in 2013 are still being prepared, but researchers told *Nature* that it is likely to be a near-record year. Across the world, almost 30 tonnes of ivory have been seized, according to events detailed in news reports and collated by TRAFFIC, a non-governmental organization in Cambridge, UK, that monitors trade in wildlife. And figures for ivory hauls in media reports collected each month by conservation group Save the Elephants, headquartered in Nairobi, add up to a similar number (see go.nature.com/4xyeln). Both numbers, however, should be regarded with caution because the size of seizures can be overestimated, and many go unreported. With each tusk providing about 5 kg of ivory, and some researchers estimating that seizures account for as little as 10% of all ivory collected, the numbers paint a bleak picture.

"I certainly don't think anything's got better this year," says Holly Dublin, chair of the elephant specialist group of the International Union for Conservation of Nature (IUCN).

Official numbers are available for 2011, when a record 46.5 tonnes of ivory was seized (see 'Tusk totals'). Samuel Wasser, director of the Center for Conservation Biology at the University of Washington, Seattle, says that poaching levels were probably higher in 2012, and that 2013 could be higher again. He estimates that around 50,000 elephants were killed in 2011, given the amount of ivory seized, and that the numbers in the two years since were similar. Figures from TRAFFIC and Save the Elephants suggest that between 25,000 and 35,000 of the animals are killed each year.

"Those numbers may be off by some margin. But based on the number of recent seizures, the elephants are being killed at their highest rate yet," says Wasser, who estimates from news reports that 38 tonnes of ivory have been seized this year.

The past year has seen an escalation of political efforts to curb poaching,

An elephant carcass in Zimbabwe, where poachers have used cyanide to kill the animals for their tusks.

PHILIMON BULAWAYO/REUTERS/CORBIS

which is increasingly being linked to large criminal syndicates and even terrorist groups. The latest such effort takes place next week in Gaborone, Botswana, under the auspices of the IUCN. African heads of state, ministers and scientists will discuss measures to fight poaching including national task forces, tougher legal action against ivory traffickers and greater use of the military against heavily armed poachers.

"We're seeing more political momentum build up," says John Scanlon, secretary-general of the Geneva-based Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). "That movement needs to be faster, but things are moving in the right direction."

At a meeting in Bangkok in March, representatives from CITES signatory countries agreed to take steps to fight the poaching scourge. These include using public-awareness campaigns to

curb demand for ivory and increased forensic tracing of seized ivory using genetic techniques.

Some positive outcomes from the CITES meeting are already being seen on the ground, says Wasser, who uses DNA analysis of seized tusks to try to trace the origin of illegal ivory by matching genetic variations across Africa. The decisions at the meeting have made "a huge difference" to the willingness of countries to provide samples, he says. Using the samples, he expects to be able to pinpoint the major hotspots of poaching, eventually enabling intensive law enforcement in those regions.

Increased political attention may already be having an effect. Nations that drive the demand for ivory are stepping up prevention efforts. Scanlon says that China, for example, is now prosecuting more people for ivory offences than in the past. And the United States — which in a show of intent earlier this month publicly crushed 6 tonnes of ivory seized at its borders since 1989, when the international ban on ivory trading was introduced — has this year set up a task force to combat the illegal wildlife trade.

Closer to the front line, George Wittemyer of Colorado State University in Fort Collins, a conservation biologist who conducts research at the Samburu National Reserve in Kenya, says that the year started with the worst poaching levels ever seen there. But he adds that killings have fallen since, driven in part by efforts to engage the local community.

"I find it relieving to see the level at which the issue is being talked about," Wittemyer says. "There are a lot of heads of state in Africa who are taking this seriously." ∎

⟳ **NATURE.COM**
For a Q&A and audio on elephant poaching, see:
go.nature.com/tpcupf

TUSK TOTALS
The amount of ivory seized worldwide reached record levels in 2011.
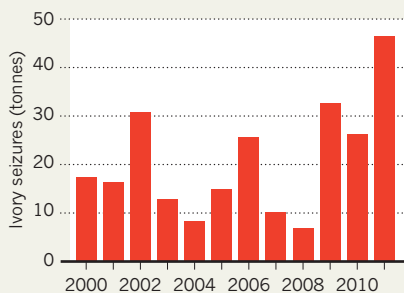
SOURCE: TRAFFIC/ELEPHANT TRADE INFO. SYSTEM/CITES

3 WAYS TO BLOW THE WHISTLE

BY ED YONG, HEIDI LEDFORD AND RICHARD VAN NOORDEN

Reporting suspicions of scientific fraud is rarely easy, but some paths are more effective than others.

Are more people doing wrong or are more people speaking up? Retractions of scientific papers have increased about tenfold during the past decade, with many studies crumbling in cases of high-profile research misconduct that ranges from plagiarism to image manipulation to outright data fabrication. When worries about somebody's work reach a critical point, it falls to a peer, supervisor, junior partner or uninvolved bystander to decide whether to keep mum or step up and blow the whistle. Doing the latter comes at significant risk, and the path is rarely simple. Some make their case and move on; others never give up. And in what seems to be a growing trend, anonymous watchdogs are airing their concerns through e-mail and public forums. Here, *Nature* profiles three markedly different stories of individuals who acted on their suspicions. Successful or otherwise, each case offers lessons for would-be tipsters.

The analytical

Uri Simonsohn sees himself as more of a data-whisperer than a whistle-blower. His day job as a social scientist at the University of Pennsylvania in Philadelphia involves scouring archival data — from house prices to auction records to college admissions — as part of his research into judgement and decision-making. He suspects that this background has predisposed him to catching spurious patterns in other psychologists' results. "With an experiment, you do a *t*-test and move on," he says. "But people who work with archival data are used to looking at data very carefully."

It was this intuition that stirred when he first came across papers by Dirk Smeesters at Erasmus University Rotterdam in the Netherlands and Lawrence Sanna at the University of Michigan in Ann Arbor in the summer of 2011. In both cases, the data seemed too good to be true, containing an overabundance of large effects and statistically significant results. In one of Sanna's papers, Simonsohn noticed that one experiment — in which volunteers were supposedly split into different groups — produced results with uncannily similar standard deviations. In the results of Smeesters' studies, he saw a suspiciously low frequency of round numbers and an unusual similarity between many of the averages. "If there's too little noise, and the data are too reliable again and again, they cannot be real," he says. "Real data are supposed to have error."

Simonsohn checked his suspicions by simulating experiments thousands of times to show how unlikely the reported results actually were. He replicated his analyses on other papers by the same authors and found the same patterns, and he carried out negative controls, showing no suspicious patterns in the work of other psychologists who used the same set-ups.

Simonsohn contacted both authors and spent months systematically ruling out alternative explanations for the discrepancies he found. Eventually, according to Simonsohn, only one remained — that they had manipulated their data. He still refrained from accusing anyone, liaising privately with Smeesters, Sanna and their co-authors, asking for raw data, outlining his concerns and asking if another party, such as a student or research assistant, could have tampered with the data. "I was extremely open-minded," he says. "My working hypothesis was that it's not in your interest to fake if you're a researcher."

Towards the end of 2011, Simonsohn learned that Erasmus University, which he had contacted, had begun an investigation. He also found out that because of his inquiries, the University of North Carolina at Chapel Hill, where Sanna had performed his work, had also started to investigate. By the summer of 2012, both Smeesters and Sanna had resigned from their posts, and several of their papers have since been retracted. In previous statements, Smeesters has said that he never fabricated data and that the practices he used are common in his field; he chose not to provide a further comment

when contacted by *Nature*. Neither Sanna nor his former institution have publicly addressed questions about his resignation and Sanna could not be reached for comment.

When asked about the two careers that have been broken by his investigations, Simonsohn pauses. "I don't feel bad about it," he concludes. "If I'm going to the same conferences as these people, and publishing in these journals, I can't just look the other way." Joe Simmons, a psychologist at the University of Pennsylvania, says that he admires his colleague's integrity and sense of obligation. "He couldn't not do something," he says.

Simonsohn hopes that his actions will spur psychologists to instigate reforms to stem fraud — one option would be to require researchers to post raw data, thereby making them more open to checks by watchful data-sleuths. He also wants researchers to disclose more details of their work at the outset of an experiment, such as the variables to analyse or their planned sample sizes. That would discourage subtler forms of data-tampering — such as continuing experiments only until results meet significance — which, in his opinion, flood the psychological literature with false positives (see *Nature* **485,** 298–300; 2012).

Simonsohn's whistle-blowing attracted its share of attention. He has received around a dozen offers to look into suspected cases of dodgy data, typically from people outside science who have personal concerns about, say, the US election. He rarely replies. He has little interest in being drawn into unnecessary disputes and bristles at any suggestion that he has led a witch-hunt — a term that he associates with the wanton use of poor diagnostic tests, not his own careful review.

"Some people think he does it for the fame, but he finds the fame annoying," says Simmons. Simonsohn, for his part, says he hopes that his new-found identity as a whistle-blower will morph into a different label, as "a person who looks carefully at data. I would be very happy with that reputation," he says.

> "IF THE DATA ARE TOO RELIABLE AGAIN AND AGAIN, THEY CANNOT BE REAL."

The quixotic

Helene Hill thought she was nearing retirement in 1999 when, one day, she decided to take a peek at a lab mate's culture dishes. A radiation biologist at the University of Medicine and Dentistry of New Jersey in Newark, Hill was collaborating with a junior colleague on a project to study the 'bystander effect', a phenomenon whereby cells exposed to an agent — in this case radiation — influence the behaviour of unexposed neighbours. Hill had trained the postdoc, Anupam Bishayee, on the technique and wanted to see how he had fared. The plates, she says, were empty, yet Bishayee later reported cell counts from them.

> "A PERSON HAS AN OBLIGATION TO DO THE RIGHT THING IF THEY CAN."

Hill would spend the next 14 years trying to expose what she believes to be a case of scientific misconduct. University panels, the US Office of Research Integrity (ORI), and two courts of law have evaluated and dismissed her concerns. Her journey has cost her thousands of dollars in legal fees and countless hours trawling through more than 30,000 documents. And it could cost her her job. Yet Hill, now 84, has no intention of backing off. "A person has an obligation to do the right thing if they can," she says.

After the first observation, Hill and another postdoc decided to covertly shadow Bishayee's experiments, snapping photos of his cultures in the incubator. When Bishayee reported data from an experiment that they thought was contaminated with mould, Hill and her colleague accused him of fabricating the results and took their concerns to the university's committee on research integrity.

But their case soon frayed. Under questioning, her colleague acknowledged that he had moved Bishayee's culture tubes before taking photos of them, which the committee viewed as potentially tampering with the evidence. And Hill explained that she had used a microscope that she was unfamiliar with when checking Bishayee's cultures. The committee determined that Hill did not have enough evidence to prove her case.

Hill would not let the matter lie. Bishayee had published his results in a paper that lists Hill as a co-author (A. Bishayee *et al. Radiat. Res.* **155,** 335–344; 2001) and his adviser, Roger Howell, used Bishayee's data to support a grant application to the National Institutes of Health (NIH) in 1999. Hill took the case to federal investigators at the ORI, who conducted a small statistical analysis of Bishayee's data. Hill says that in her opinion the patterns therein suggested fabrication, and one ORI investigator, Kay Fields, thought the case had merit. But Fields was overruled by a superior, in part because he believed that the control data for the analysis — Hill's own — were also statistically questionable. The ORI determined that there was insufficient evidence to prove misconduct.

Hill continued to petition her university and the ORI to review the data. Fields, meanwhile, says that she felt obliged to tell Hill about another option: a '*qui tam*' lawsuit. Such lawsuits, allowed under the US False Claims Act, can be brought by any citizen to aid the government in recouping taxpayers' funds allocated under false pretences. Hill's case could be eligible because of the NIH grant.

Qui tam can be a risky strategy, says David Lewis, director of the research misconduct project at the nonprofit National Whistleblower Center in Washington DC. He has filed two *qui tam* lawsuits in the past, unrelated to Hill's (see *Nature* **453,** 262–263; 2008). Both were unsuccessful, and Lewis generally doesn't recommend the strategy. In Hill's case, the process dragged on for years and cost her US$200,000 in legal fees. "I don't think my children are too happy with my having lost that much money," she says, "but I just felt I had an obligation to see it through."

New Jersey District Court judge Dennis Cavanaugh ruled in favour of Bishayee and Howell in October 2010, and referred to Hill's battle as "a quest of Quixotic proportions that ultimately must be put to rest". Hill lost her final appeal in October 2011. Still, she says that her investment paid off: the discovery phase of the lawsuit allowed her access to ten years' worth of the Howell lab's notebooks.

With those data in hand, she teamed up with statistician Joel Pitt of Georgian Court University in Lakewood Township, New Jersey. Together, they pored over data that Bishayee had hand-recorded from a machine that counts cells. The duo also gathered larger control data sets from others who had used the same machine. Pitt looked at the frequency of the numbers appearing as the least significant digit of each recorded count. These should have a random distribution, but Bishayee's data seemed to favour certain numbers. Pitt calculated the odds of those frequencies arising by chance as less than 1 in 100 billion. In Hill's view, the implication is clear: Bishayee made the numbers up.

Along with Pitt, Hill has been trying, so far unsuccessfully, to publish these statistical analyses and further publicize her allegations, actions that Robert Johnson, the dean of her institution — now part of Rutgers University — warned in a strongly worded letter in July could lead to "additional disciplinary action, up to and including termination".

Howell, in a written statement to *Nature*, expressed frustration at the time spent revisiting the issue despite no finding of wrongdoing. Bishayee did not respond to *Nature*'s request for comment. Fields says: "I admire Dr Hill for the courage of her convictions, but it is difficult to say that she was prudent to pursue the case for so long and at such expense."

Hill, for her part, remains undeterred. "I want to finish," she says. "It becomes almost an obsession."

The anonymous

Anonymous tipsters are nothing new. But since 2010, someone going by the pseudonym 'Clare Francis' has seriously upped the ante. She or he (or they; many suspect it is a group of people) has sent hundreds of e-mails to life-science journal editors, flagging up suspected cases of plagiarism or instances in which figures appear to be manipulated or duplicated. Her terse, sometimes cryptic complaints have resulted in a handful of retractions and corrections, but editors have felt bombarded by her voluminous notices — many of which, they say, lead nowhere.

Like her or not, Francis has sparked a debate about how editors deal with anonymous tips, which are now poised to grow thanks to the proliferation of websites that allow anyone to publicly air grievances about research papers.

Sabine Kleinert, a senior executive editor at *The Lancet* and former vice-chair of the UK-based Committee on Publication Ethics (COPE), calls the recent surge in anonymous comments "the Clare Francis phenomenon". Phenomenon is an apt descriptor. Francis estimates that she has e-mailed "about 100" different editors. And those publishers who agreed to talk to *Nature* said that their editors generally receive multiple messages from her. Diane Sullenberger, executive editor of the *Proceedings of the National Academy of Sciences*, says that as many as 80% of the allegations they receive come from Francis. And the scientific publisher Wiley says that in 2011 Francis's name was on more than half of its investigation requests.

Anonymity generally makes people uncomfortable, says Ulrich Brandt, editor-in-chief of *Biochimica et Biophysica Acta*. "One has to wonder about the motivation of the whistle-blower," he says. "Ill-founded allegations of scientific misconduct can do harm and may constitute a form of scientific misconduct themselves."

By 2011, editors were growing increasingly frustrated by Francis because — quite apart from her anonymity — many of her claims did not check out. "I have no problem taking time to look at an allegation — but I don't like people wasting my time," says Eric Murphy, editor-in-chief of *Lipids*. Moreover, many of Francis's complaints are oblique and hard to follow, says Sullenberger. "It is helpful to know specific details about the concerns from a scientific standpoint, not just, 'The bands in the 10- and 60-minute lanes are geometrical and superimposable' or 'Background is silvery smooth,'" she says, referring to some of Francis's e-mails.

Some journal editors have warned Francis that they are less likely to follow up on her requests than other complaints. In September 2011, Wiley's then legal director, Roy Kaufman, sent her an e-mail saying that the company could "not guarantee that all anonymous allegations sent to us will be investigated". Francis made the note public, sparking debate over how such allegations should be handled.

Two years on, the attitudes of editors have changed to some degree. In February this year, COPE put out guidelines on "responding to anonymous whistle blowers". Francis was not mentioned by name, but was the main driving force behind the work, says Virginia Barbour, COPE's current chair. "Editors were feeling guilty, and upset, and didn't understand how they should approach it." COPE reminded them that, no matter where they came from, "all allegations … that have specific, detailed evidence to support the claim should be investigated". But Anna Trudgett, editorial director at the journal *Blood*, says that the journal still addresses Francis's e-mails only selectively. "Not all anonymous correspondence is treated the same way," she says. Wiley has adjusted its practice to investigate all complaints, says spokesperson Helen Bray.

Fundamentally, editors are not just reacting to Clare Francis's pseudonymity. They are also irritated by the way she works. "For some, it's not that Clare Francis is a pseudonym; it's that the pseudonym is Clare Francis," says Tom Reller, a spokesperson for Elsevier. Some editors bring up what they say is Francis's aggressive tone and pursuit of lost causes. "When we determine that the allegation is not founded, it is not uncommon for Clare Francis not to accept the result," says Véronique Kiermer, Nature Publishing Group's executive editor.

In Barbour's view, Francis's tactics are not a good model for other anonymous tipsters to emulate. To make up for the inevitable loss of trust that comes from being anonymous, tip-offs gain credence if they are precise, detailed and polite. Francis sometimes meets these standards but often does not.

To Francis, such critiques miss the point. Asked about her tone, she wrote back: "I do not have a 'tone'. I try to describe what I can see." She adds that editors often focus narrowly on their journal when she sends what she says are connected patterns of image manipulation across many journals. "They will not look at the whole picture, but remain in purdah," she writes. As for alleged false leads, she says: "The hit rate would be higher if they paid attention to what is on the page rather than their fantasy world."

One thing that editors and Francis might agree on is that anonymous whistle-blowing is likely to increase, given the increased access to papers by people all around the world and the availability of online tools for spotting potential plagiarism and image manipulation. One site, called PubPeer, is already becoming a venue for anonymous comments — including postings in a similar vein to Francis's style.

The growth here is a sign that whistle-blowers are not being protected enough within the academic environment, says Kleinert. "This is where we have to do much more. Somebody should feel comfortable to be able to raise issues without fearing retaliation or damage to their own career". ∎

"I DO NOT HAVE A 'TONE'. I TRY TO DESCRIBE WHAT I CAN SEE."

Ed Yong *is a science journalist in London;* **Heidi Ledford** *writes for* Nature *from Cambridge, Massachusetts; and* **Richard Van Noorden** *writes for* Nature *from London.*

The pursuit of happiness

Researchers have struggled to identify how certain states of mind influence physical health. One biologist thinks he has an answer.

BY JO MARCHANT

When Steve Cole was a postdoc, he had an unusual hobby: matching art buyers with artists that they might like. The task made looking at art, something he had always loved, even more enjoyable. "There was an extra layer of purpose. I loved the ability to help artists I thought were great to find an appreciative audience," he says.

At the time, it was nothing more than a quirky sideline. But his latest findings have caused Cole — now a professor at the Cousins Center for Psychoneuroimmunology at the University of California, Los Angeles — to wonder whether the exhilaration and sense of purpose that he felt during that period might have done more than help him to find homes for unloved pieces of art. It might have benefited his immune system too.

At one time, most self-respecting molecular biologists would have scoffed at the idea. Today, evidence from many studies suggests that mental states such as stress can influence health. Still, it has proved difficult to explain how this happens at the molecular level — how subjective moods connect with the vastly complex physiology of the nervous and immune systems. The field that searches for these explanations, known as psychoneuroimmunology (PNI), is often criticized as lacking rigour. Cole's stated aim is to fix that, and his tool of choice is genome-wide transcriptional analysis: looking at broad patterns of gene expression in cells. "My job is to be a hard-core tracker," he says. "How do these mental states get out into the rest of the body?"

With his colleagues, Cole has published a string of studies suggesting that negative mental states such as stress and loneliness guide immune responses by driving broad programs of gene expression, shaping our ability to fight disease. If he is right, the way people see the world could affect everything from their risk of chronic illnesses such as diabetes and heart disease to the progression of conditions such as HIV and cancer. Now Cole has switched tack, moving from negative moods into the even more murky territory of happiness. It is a risky strategy; his work has already been criticized as wishful thinking and moralizing. But the pay-off is nothing less than finding a healthier way to live.

"If you talk to any high-quality neurobiologist or immunologist about PNI, it will invariably generate a little snicker," says Stephen Smale, an immunologist at the University of California, Los Angeles, who is not affiliated with the Cousins Center. "But this doesn't mean the topic should be ignored forever. Someday we need to confront it and try to understand how the immune system and nervous system interact."

THE BEST MEDICINE?

In 1964, magazine editor Norman Cousins was diagnosed with ankylosing spondylitis, a life-threatening autoimmune disease, and given a 1 in 500 chance of recovery. Cousins rejected his doctors' prognosis and embarked on his own programme of happiness therapy, including regular doses of Marx Brothers films, and credited it with triggering a dramatic recovery. He later established the Cousins Center, which is dedicated to investigating whether psychological factors really can keep people healthy.

At the time, mainstream science rejected the idea that any psychological state, positive or negative, could affect physical well-being. But studies during the 1980s and early 1990s revealed that the brain is directly wired to the immune system — portions of the nervous system connect with immune-related organs such as the thymus and bone marrow, and immune cells have receptors for neurotransmitters, suggesting that there is crosstalk.

> ## "Mood matters. If we change the psychology, physiological changes do parallel that."

These connections seem to have clinical relevance, at least in the case of stress. One of the first researchers to show this was virologist Ronald Glaser, now director of the Institute for Behavioral Medicine Research at the Ohio State University in Columbus. "When I started working on this in the 1980s, nobody believed what stress could do, including me," he recalls. He and his colleagues sampled blood from medical students, and found that during a stressful exam period, they had lower activity from virus-fighting immune cells[1], and higher levels of antibodies for the common virus Epstein–Barr[2], suggesting that stress had compromised their immune systems and allowed the normally latent virus to become reactivated.

The field of PNI has grown hugely since then, with medical schools worldwide boasting their own departments of mind–body medicine, of which PNI is just one component. It is now accepted that the body's response to stress can suppress parts of the immune system and, over the long term, lead to damaging levels of inflammation. Large epidemiological studies — including the Whitehall studies, which have been following thousands of British civil servants since 1967 — suggest[3] that chronic work stress increases the risk of coronary heart disease and type 2 diabetes, for example. Low socio-economic status increases susceptibility to a wide range of infectious diseases, and there is considerable evidence that stress increases the rate of progression of HIV/AIDS. But researchers have a long way to go before they will understand exactly how signals from the brain feed into physical health.

WORRIED SICK

PNI studies have mostly tended to look at levels of individual immune-cell types or molecular messengers — such as the stress hormone cortisol and the immune messenger proteins called cytokines — or the expression of individual genes. But Cole wanted to get a sense of how the whole system was working.

His first foray, published in 2007, looked at loneliness[4]. Social isolation is one of the most powerful known psychological risk factors for poor health, but it is never certain whether it causes the health problems, or whether a third factor is involved: lonely people might be less likely than others to eat well, for example, or to visit their doctor regularly.

Cole and his colleagues looked at gene expression in the white blood cells of six chronically lonely people — people who had said consistently over several years that they felt lonely or isolated, and were fearful of other people — and eight people who said that they had great friends and social support. Out of the roughly 22,000 genes in the human genome, the researchers identified 209 that distinguished the lonely people from the sociable ones: they were either regulated up to produce

A volunteer helps to bag meals for the homeless at Cathedral Kitchen in Camden, New Jersey.

more of an individual protein or regulated down to produce less. Any individual gene could easily look different by chance, but Cole was struck by the overall pattern. A particularly large proportion of the upregulated genes in the lonely group turned out to be involved in the inflammatory response, whereas many of the downregulated genes had antiviral roles. In sociable people, the reverse was true. It was a small study, but one of the first to link a psychological risk factor with a broad underlying change in gene expression.

The researchers have since replicated that result in a group of 93 people[5]. Cole says that he has also seen a similar shift in gene expression in individuals exposed to various types of social adversity, from imminent bereavement to low socio-economic status.

The results make evolutionary sense, he says. Early humans in close-knit social groups would have faced increased risk of viral infections, so they would have benefited from revved-up antiviral genes. By contrast, people who were isolated and under stress faced greater risk of injuries that could cause bacterial infection — and thus would need to respond by ramping up genes associated with inflammation, to help heal wounds and fight off those infections. But modern stresses lead to chronic and unhelpful inflammation, which over time damages the body's tissues, increasing the risk of chronic diseases such as atherosclerosis, cancer and diabetes.

To a classical immunologist such as Smale,

Cole's results are "intriguing, wonderful observations", but not yet completely convincing. In future work, he wants to see the rest of the physiological pathway nailed down. "Until you put together a full understanding of that mechanism, you have this level of uncertainty and scepticism," he says. That sentiment is echoed by Alexander Tarakhovsky, an immunologist at the Rockefeller University in New York City. Pinning down precise mechanisms — for example, which neurotransmitters cause which specific effects — is extremely difficult, he says, because the brain and the immune system are both so complex. Cole's research "makes you think about what the consequences of social hardship could be, but it doesn't really tell you how it works".

Greg Gibson, director of the Center for Integrative Genomics at the Georgia Institute of Technology in Atlanta, wants to see larger studies but argues that the big-picture "genetic architecture" that Cole is uncovering is worth studying, even if not every detail of the mechanism is yet understood. "A lot of people are taking a whole-genome approach, but they focus only on a handful of 'top hits'. They are missing the wood for the trees."

DON'T WORRY, BE HAPPY

In 2010, Cole received an e-mail from Barbara Fredrickson, a friend from graduate school who was now studying emotional well-being at the University of North Carolina in Chapel Hill. "Remember me?" she said. She

was interested in the biological correlates of happiness and other positive emotional states, and suggested that the pair collaborate. After years of looking at stress and adversity, Cole loved the idea. "I was bored as hell with misery," he says.

If PNI as a whole has credibility issues, studying well-being is even trickier. It is more slippery to measure than stress — there is no biological marker such as cortisol to fall back on and no simple way to induce it in the lab, and mainstream biologists tend to look down on fuzzy methods of data collection such as questionnaires.

One approach is to test whether it is possible to reverse the adverse effects on gene expression caused by stress. Cole has collaborated in three small, randomized, controlled trials that attempt to do this. Studies involving 45 stressed caregivers[6] and 40 lonely adults[7] respectively found that courses in meditation shifted gene-expression profiles in the participants' white blood cells away from inflammatory genes and towards antiviral genes. A third trial[8], led by psycho-oncologist Michael Antoni at the University of Miami, Florida, involved 200 women with early-stage breast cancer. In those who completed a ten-week stress-management programme, genes associated with inflammation and metastasis were downregulated compared with those of women in the control group, who attended a one-day educational seminar. Meanwhile, genes involved in the type I interferon response

Psychoneuroimmunologist Steve Cole studies how stress and happiness affect health.

(which fights tumours as well as viruses) were upregulated in the women who took the stress-management course. "Our conclusion was that mood matters," says Antoni. "If we change the psychology, physiological changes do parallel that."

Cole and Fredrickson aspired to go further. Instead of looking at the benefits of blocking stress, they wanted to investigate what happens in the body when people are happy. To that end, they asked 80 participants 14 questions, such as how often in the past week they had felt happy or satisfied, and how often they felt that their life had a sense of meaning[9]. The questions were designed to distinguish between the two forms of happiness recognized by psychologists: hedonic well-being (characterized by material or bodily pleasures such as eating well or having sex) and eudaimonic well-being (deeper satisfaction from activities with a greater meaning or purpose, such as intellectual pursuits, social relationships or charity work).

The researchers were surprised to find that the two types of happiness influenced gene expression in different ways. People with a meaning-based or purpose-based outlook had favourable gene-expression profiles, whereas hedonic well-being, when it occurred on its own, was associated with profiles similar to those seen in individuals facing adversity.

One interpretation is that eudaimonic well-being benefits immune function directly. But Cole prefers to explain it in terms of response to stress. If someone is driven purely by hollow consumption, he argues, all of their happiness depends on their personal circumstances. If they run into adversity, they may become very stressed. But if they care about things beyond themselves — community, politics, art — then everyday stresses will perhaps be

of less concern. Eudaimonia, in other words, may help to buffer our sense of threat or uncertainty, potentially improving our health. "It's fine to invest in yourself," says Cole, "as long as you invest in lots of other things as well."

PERILS OF POSITIVE THINKING
This is just the kind of advice that attracts some of the most vociferous criticisms of Cole's work. James Coyne, a health psychologist and emeritus professor at the University of Pennsylvania in Philadelphia, says that Cole and Frederickson's well-being study is simply too small to show anything useful. He also argues that the measures of eudaimonic and hedonic happiness are so highly correlated in the study as to be essentially the same thing. Coyne says that early results are being vastly over-sold. "They claim that if you make the right choices, you'll be healthy. And if you don't, you'll die."

Coyne wants researchers across the field of PNI to stop publicizing claims about health benefits until the science is more solid. "They're turning it into books and workshops, telling people how to live their lives."

Fredrickson, for example, is the author of two popular books, including *Positivity* (Crown Archetype, 2009), which posits that a specific ratio of positive to negative emotions (2.9013, to be precise) is linked to good health. The book has been praised by eminent psychologists such as Daniel Goleman and Martin Seligman, but the set of equations behind the ratio was criticized this year[10] by Alan Sokal, a physicist at New York University (who famously published a deliberately nonsensical paper in the journal *Social Text* in 1996, intended to expose the lack of rigour in the field of cultural studies). He pointed out that the equations are based on parameters from a 1962 paper on air flow, with no

connection to psychological data at all. Fredrickson acknowledges problems with the maths, which she based on a peer-reviewed paper on the complex dynamics of teams[11], but says that she stands by the fundamental principles described in the book. "There seems good enough evidence to suggest that emotions contribute to health."

Cole and Fredrickson agree that their study is small and needs to be repeated. But they say that extensive previous research has validated the questionnaire they used and confirmed that it measures two distinct, albeit highly correlated, emotional states. They also note that correlation does not necessarily mean that two states are the same: height and weight are also highly correlated, for example, yet describe different things. Each type of happiness tends to encourage the other, says Fredrickson, "but we can try to understand which is leading the way towards health".

The researchers are not the first from the PNI community to face accusations of wishful thinking. Indeed, the story of the field's founder — hailed in the press as proof of the power of positive emotions — has been questioned. Immunologists have suggested that Cousins was not suffering from ankylosing spondylitis at all, but from polymyalgia rheumatica, which often clears up on its own. His "health probably coincidentally remitted", says Cole.

Despite the criticisms, and the fact that his work is in its early days, Cole says that he is struck by the evidence that positive emotions can override the biological effects of adversity — enough to make changes in his own life. Although he no longer has time to engage in the art trade, he has embraced the ways that his hobby helped him. "I have spent most of my career and personal life trying to avoid or overcome bad things," he says. "I spend a lot more time now thinking about what I really want to do with my life, and where I'd like to go with whatever years remain." ∎

Jo Marchant *is a freelance science journalist based in London.*

1. Kiecolt-Glaser, J. K. *et al. Psychosom. Med.* **46,** 7–14 (1984).
2. Kiecolt-Glaser, J. K., Speicher, C. E., Holliday, J. E. & Glaser, R. *J. Behav. Med.* **7,** 1–12 (1984).
3. Cohen. S., Janicki-Deverts, D. & Miller, G. E. *J. Am. Med. Assoc.* **298,** 1685–1687 (2007).
4. Cole, S. W. *et al. Genome Biol.* **8,** R189 (2007).
5. Cole, S. W., Hawkley, L. C., Arevalo, J. M. G. & Cacioppo, J. T. *Proc. Natl Acad. Sci. USA* **108,** 3080–3085 (2011).
6. Black, D. S. *et al. Psychoneuroendocrinology* **38,** 348–355 (2012).
7. Creswell, J. D. *et al. Brain Behav. Immun.* **26,** 1095–1101 (2012).
8. Antoni, M. H. *et al. Biol. Psychiatry* **71,** 366–372 (2012).
9. Fredrickson, B. L. *et al. Proc. Natl Acad. Sci. USA* **110,** 13684–13689 (2013).
10. Brown, N. J. L., Sokal, A. D. & Friedman, H. L. *Am. Psychol.* http://dx.doi.org/10.1037/a0032850 (2013).
11. Losada, M. *Math. Comput. Model.* **30,** 179–192 (1999).

COMMENT

Moldovan police examine suspected radioactive uranium-238 in August 2010.

Expand nuclear forensics

Characterizing nuclear materials deters illicit trafficking and terrorism, but more scientists, techniques and collaborations are needed, says **Klaus Mayer**.

Since the International Atomic Energy Agency (IAEA) implemented its Incident and Trafficking Database in 1995, around 2,300 events involving illicit nuclear or other radioactive materials have been reported. Although most cases involve lost or orphan radioactive sources containing, for example, cobalt-60 or iridium-192 for medical or industrial applications, 10–15 incidents per year concern nuclear materials turning up out of regulatory control.

Uranium and plutonium are most worrying because, as well as posing a radiological hazard, they may be indicative of proliferation or nuclear terrorism. The sorts of things seized are scrap metal contaminated with grams of enriched uranium or kilograms of natural uranium, gram-sized samples of uranium metal, and uranium fuel pellets. In 1994, 300 grams of plutonium oxide powder were intercepted at Munich airport in Germany.

Officials detect unlawful nuclear materials at borders, seaports and airports or in state territories by measuring radiation directly or acting on tip-offs from police or intelligence work. Whenever such a sample is intercepted, agencies want to know: which laws have been broken? When and where was the material produced? What was the intended use? Where was the material stolen or diverted? Is more of it at large? Nuclear-forensic scientists try to answer these questions.

The chemical and physical signatures of a radioactive material — from its appearance and microstructure to its elemental and isotopic composition — shed light on its origin and history. For example, the isotope ratios of strontium impurities in a sample of natural uranium may indicate whether it was mined in Australia or Namibia. The presence of daughter products from nuclear decays reveal the production date of the material, and products, such as uranium-236, of neutron reactions indicate that it was irradiated in a power plant.

Nuclear forensics is a small and specialized field that has matured since the early 1990s. But progress is still too slow. Although the number of scientific publications in the discipline has risen from a handful in 2001, it still numbers only a few dozen a year.

States worldwide need to implement nuclear-forensic capabilities — both nationally and internationally — through greater collaboration. To boost the robustness of the methods, and thence their credibility, new forms of analysis and signatures for nuclear materials need to be developed. Nuclear-forensic data need to be archived securely and more experts must be trained. Otherwise smugglers and terrorists might evade prosecution.

A few years ago in a European country, a radiation detector at a scrap-metal recycling facility triggered an alarm. A piece of steel in a shipment from south Asia had ▶

▶ a greenish deposit that a rapid on-site measurement showed was natural uranium.

A sample was sent to our nuclear-research laboratory in Karlsruhe, Germany, where my team and I identified the green material as uranium tetrafluoride, an intermediate product of uranium processing encountered typically during isotope enrichment. Dating suggested that it was produced in 1978. But chemical impurities, in particular the pattern of the rare-Earth elements (including lanthanum, neodymium and samarium), indicated that the uranium came from a sandstone subtype found not in the suspected country of origin, but in China, Australia, Niger or the Czech Republic (see 'Chemical impurities').

Uranium production in the suspected country had, according to the literature, started in the late 1970s from a different type of sandstone. Further literature studies revealed that in the same period, uranium-ore concentrate from Niger, fitting the seized sample's characteristics, was imported into the suspected country. Thus, the origin and history of the material showed that uranium processing and isotopic enrichment had already been achieved at a very early stage in the country's nuclear activities.

NUCLEAR FINGERPRINT

Chemical and physical signatures vary through the nuclear fuel cycle: from uranium ore to uranium fuel pellets used in power reactors, natural uranium to weapons-usable, highly enriched material, and spent nuclear fuel to separated plutonium. The variety of materials reflects the diverse geological and geographic origins of natural uranium, and the technological processes that could have been applied add to the diversity.

Analysis methods must be tailored to the material and signature under investigation. Uranium and plutonium samples, for instance, contain different chemical impurities that require different treatments.

Nuclear-forensic interpretations build on a variety of measurements — including mass spectrometry, electron microscopy, α or γ spectrometry and radiochemical separations — that yield a broad spectrum of material parameters. These range from obvious characteristics such as uranium enrichment or pellet dimensions to more sophisticated information including metallic impurities or grain-size distribution.

Age is derived from ratios of progenies of radioactive parent nuclides. Most other signatures are comparative, referenced against samples of known provenance. Some reference data, such as the rare-Earth-element

patterns studied by isotope geologists, are openly published. Information such as the grain-size distribution in uranium fuel pellets can be provided only by the producer.

There are broader challenges in nuclear forensics: new analytical methods need to be validated, the robustness of some signatures needs to be demonstrated and the interpretational techniques need to be substantiated.

CHEMICAL IMPURITIES

The signature of rare-Earth elements in natural uranium differs depending on the type of ore from which the uranium is mined. The concentration profile of a seized sample of nuclear material matches the sandstone type of a mine in Arlit, Niger, corroborating information that the material originates from the country.
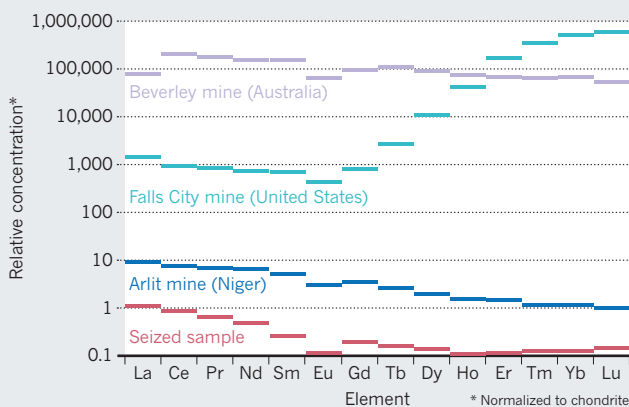
There is no centralized international nuclear-forensic database. Indeed, it is fiercely resisted by many nations, for understandable reasons. Data characterizing nuclear materials and processing histories are sensitive and may be classified. Sensitivities can be commercial (in the case of nuclear fuels) or security-related (for weapons-grade uranium or plutonium). Any compilation of data on nuclear material must be secure.

A decentralized approach is gaining acceptance. The concept of national nuclear-forensic libraries, a combination of databases and physical sample archives that allows states to control their own nuclear-materials data, is being promoted by the IAEA, headquartered in Vienna, and the Global Initiative to Combat Nuclear Terrorism (GICNT). Although few countries have taken official steps, Ukraine is developing such a library, as are some others in the European Union and southeast Asia. Comparing signatures of seized material against stored information will reveal whether the material is of domestic origin. Private queries to other states could help to identify the legal owner of the material in confidence.

Yet at the same time, skilled radiochemists, nuclear physicists and nuclear engineers with hands-on experience in the nuclear fuel cycle and in production or analysis of nuclear material have become a rarity, as a report by the American Physical Society and the American Association for the

Advancement of Science highlighted (see go.nature.com/ckflpx).

Capacity building in nuclear forensics is the key issue in forming a global response to illicit trafficking and nuclear terrorism. Effective deterrence does not necessarily imply investing enormous budgets to establish sophisticated laboratories. Measurements of a few parameters may provide enough information for law-enforcement purposes. The isotopic composition of uranium or plutonium, for instance, can be determined using portable γ-spectrometry instruments, which cost about US$130,000.

GROW EXPERTISE

Building a nuclear-forensic workforce requires a scientific education in chemistry or physics with specialization in radiochemistry, nuclear engineering or nuclear physics. Hands-on experience working with nuclear material and analytical techniques comes next. Opportunities for graduate and postgraduate students to specialize in nuclear forensics should be offered through university courses and through internships and placements in nuclear laboratories.
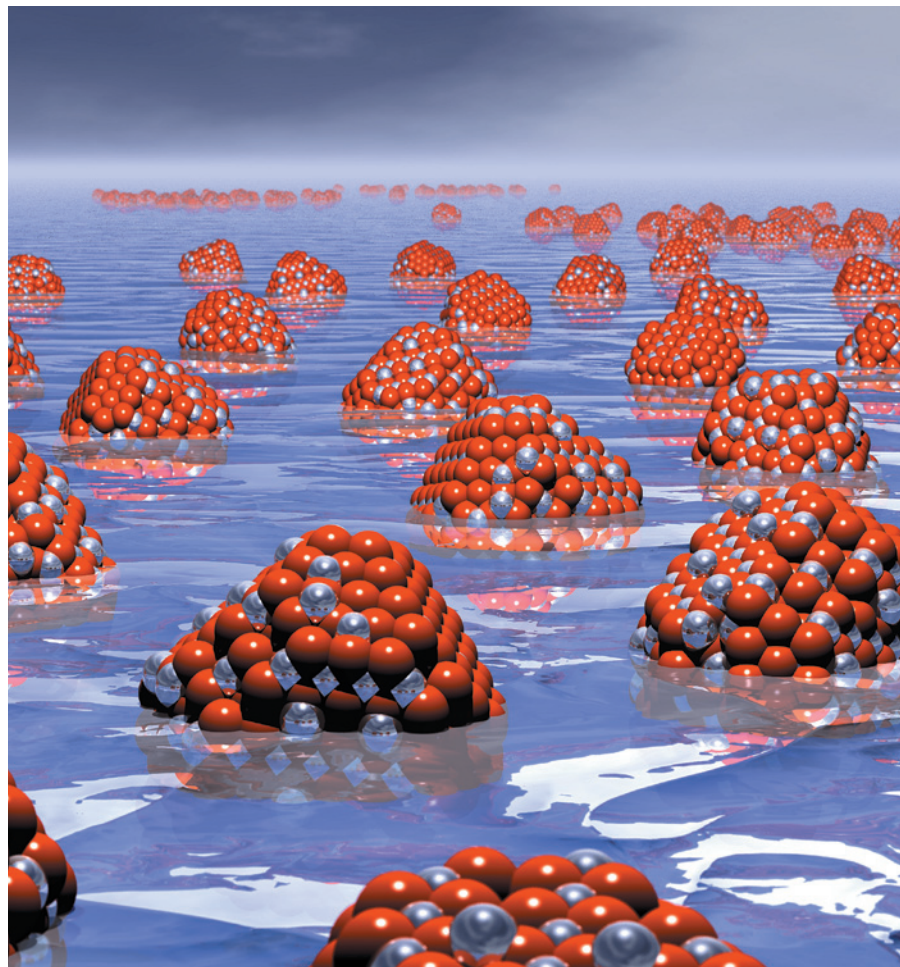
Training programmes should be harmonized and coordinated around the world. National and international exercises would demonstrate and develop competencies, and check interagency cooperation and levels of preparedness. Curricula and materials could be reviewed by the Nuclear Forensics International Technical Working Group — a gathering of nuclear forensics practitioners, including scientists and law enforcers, founded in 1995 on the initiative of the Group of Eight (G8) countries.

Mechanisms need to be developed to ensure the security and sharing of information about nuclear materials held in national databases. The GICNT should promote national nuclear-forensic libraries. The IAEA, with its expertise in assisting states with nuclear security, is well positioned to provide technical guidance.

In March 2014, the international Nuclear Security Summit will try to enhance international cooperation to prevent malicious use of nuclear material. These discussions must call on the 53 participating nations to increase awareness of the opportunities that nuclear-forensic science offers to ensure nuclear security around the globe. ∎

Klaus Mayer *is Action Leader for Forensic Analysis and Combating Illicit Trafficking at the European Commission Joint Research Centre, Institute for Transuranium Elements, Karlsruhe, Germany.*
e-mail: klaus.mayer@ec.europa.eu

existing collaborations in nanotechnology. The National Nanotechnology Initiative (NNI), established a decade earlier for materials in the 1–100-nanometre range, is a ready partner for the MGI, which encompasses scales from nanometres to micrometres.

The MGI could consider joining the NNI's Nanotechnology Knowledge Infrastructure initiative that was launched in May 2012 to develop a digital data and information framework and to strengthen collaborations between the science and modelling communities.

> *"The MGI can serve as a hub for sharing information."*

This initiative has already defined a set of Data Readiness Levels, modelled on NASA's Technology Readiness Levels, to provide a basis for communicating the quality and maturity of materials data.

The MGI could also join the partnership between the NNI and the European Commission to support a transatlantic dialogue on the nuts and bolts of data sharing: informatics, consensus-derived ontologies, data representation and archiving.

Data sharing is an inherently collaborative activity that has the potential to propel materials science forward more rapidly. The MGI can invigorate existing efforts and serve as a hub for sharing information on materials at all scales.

Sharing data in materials science

Two years on from the launch of the US Materials Genome Initiative, five experts highlight how materials scientists still need to work differently.

SALLY TINKLE
Learn from other initiatives

Policy analyst at the Science and Technology Policy Institute, Washington DC

The US Materials Genome Initiative (MGI), launched in June 2011 by President Barack Obama, aims to halve the time and cost of developing advanced materials for applications such as energy, transport and security. Two years in, hundreds of millions of dollars have been invested in academic, industry and federal-agency projects.

Sharing data and developing computational tools are crucial to the MGI's success. Advanced materials have complex physical and chemical properties that can be manipulated for different applications, and these can change during synthesis, manufacture and use. The tracking of these properties is a formidable task, and the MGI includes efforts to standardize terminology, data-archiving formats and reporting guidelines.

Fortunately, much can be learned from

DAVID L. MCDOWELL
Incentivize sharing

Executive director of the Institute for Materials, Georgia Institute of Technology, Atlanta

The MGI must avoid a 'build it and they will come' attitude. Incentives are needed for scientists and engineers to collaborate and share their data and skills. There has to be something in it for everyone.

The data-sharing environment must invite collaboration as well as facilitate it. Stakeholders have broad interests that go beyond retrieving existing data — they want to discover materials and forecast enhanced products. An intuitive and robust online environment, and cyber-infrastructure growth that is distributed and organic, rather than centralized, will encourage contributions from diverse users.

Social-networking strategies can connect users with varied expertise to pursue common interests. Win–win approaches should be encouraged. For example, uploading experimental data sets in return

for access to modelling tools drives further modelling. Clear agreements must govern credit attribution and the ethics of data use.

Maximizing the utility of information is a major attraction for investors in the MGI's infrastructure. Expensive data sets obtained, for example, from national synchrotron and neutron-diffraction facilities should be archived and leveraged to the greatest extent possible for searching and citation, as should data from massive supercomputer simulations.

Open-access rules are desirable, following the example of the National Science Foundation-funded nanoHUB for nanometre-scale modelling and simulation tools, as well as the LAMMPS molecular-dynamics code and the DREAM.3D software for meshing three-dimensional microstructures.

AMANDA BARNARD
Embrace uncertainty

Head of the Virtual Nanoscience Laboratory, Commonwealth Scientific and Industrial Research Organisation, Parkville, Australia

The MGI is opening up styles of collaborative working that raise technological and personal challenges. Materials scientists must become more comfortable with uncertainty. They must relinquish control, trust their fellow scientists, and resist the urge to redo everything 'just to be sure'.

Delivering new science from existing data requires the pooling of resources. Some insights and breakthroughs cannot be made any other way. One method may probe scales or achieve resolutions that others cannot. Electron microscopy can resolve subatomic features on surfaces, but optical microscopy shows how light reflects from them.

It is difficult to combine results from different sources. Errors arise from idiosyncrasies in experimental or computational techniques. Many experimentalists know the frustration of reproducing results that vary with laboratory conditions. Even theory-based computational methods can yield different answers.

Mixing data from different origins often introduces more uncertainty than a simple sum of the measurement or statistical errors stemming from the pure data sets. To benefit from data sharing, we must learn to live with that.

The other sort of uncertainty that MGI users must embrace is the human element — our opinions of the people who created the original data and of their competence.

Scientists are trained to be sceptical as well as objective. To move materials research forward quickly, we need to assume that each contributor is highly capable, and let the quality of the data speak for itself.

The MGI's value will only come if we can draw from it as easily and confidently as we give to it.

FRANCOIS GYGI
Make simulations reproducible

Professor of computer science, University of California, Davis

The most rapid rewards of the MGI could come from sharing simulations of materials structures.

Numerical simulations are not as reliable and reproducible as their theoretical and computational basis would suggest. They often give differing results owing to the complexity of approximations and the number of parameters used.

Overcoming these difficulties is essential for designing new materials. More robust predictions from simulations of the formation of defects in the lattice of a material, for example, improves our ability to optimize the materials' strength or electronic properties.

"Universal data formats and centralized databases are not always necessary."

Data are reliable only if they can be independently verified and reproduced by different research groups, ideally using different tools. Sharing data freely will make such cross-validation possible.

When disseminating simulation data, researchers must bear two points in mind. First, simulation software should be openly accessible, not just results. Software vendors must not forbid — as some currently do — publication of raw results or performance data out of fear that comparisons may show their product in an unfavourable light. The scientific community should fight this trend.

Second, universal data formats and centralized databases are not always necessary. The materials community could adopt existing frameworks for data sharing. For example, a vast amount of open-source software already supports the World Wide Web Consortium standards for publishing and exchanging data on the Internet, such as the Extensible Markup Language (XML).

With a modest investment, researchers can publish their own data on their own servers in ways that others can access

readily. By encouraging the development of domain-specific web tools, we will lower the barriers to data cross-verification and validation.

PETER B. LITTLEWOOD
Probe the infinite variety

Associate laboratory director for physical sciences and engineering, Argonne National Laboratory, Illinois

From synchrotrons to scanning-electron microscopes, nanotechnology tools have been honed in the information revolution. Now, through the MGI, we need to invent molecular manufacturing by expanding our vision to include the infinite variety of materials.

There are fundamental hurdles. Despite the initiative's ambitious name, atoms are not genes. The biological genome is both a theory and an algorithm for execution. In materials science, quantum mechanics can doom attempts to translate perfectly from code to function.

This theoretical impasse simply reflects the diversity of materials. Tiny variations in composition or structure can produce entirely new functions. The semiconductor industry depends on a delicate salting of silicon with minute concentrations of other atoms.

Yet chemistry can be systematic. Since Dmitri Mendeleev formulated the periodic table, we have exposed patterns of materials' structure and function, now sifted with the aid of powerful computers and high-throughput experiments. We are building, if not a single 'genome', a patchwork of tools matched to material type, property and function. The MGI will expand that.

But the brute-force approach of the modern electronics industry cannot be scaled up to make lightweight structural materials, batteries or solar cells. Here, production must be measured in megatonnes and square kilometres. The MGI has to help us beyond design and into synthesis — our goal being the engineering of programmable matter that builds itself. ∎

CORRECTION

The Comment 'Melting glaciers bring energy uncertainty' (*Nature* **502,** 617–618; 2013) wrongly said that Himalayan glaciers lost 174 gigatonnes of water each year for the period 2003–09. This was not the annual rate, but the total amount for that period. And the Indus depends on glacial waters for up to half of its flow, not half of its flow, as stated.

A forensics specialist from the International Commission on Missing Persons examines human remains from a mass grave in Tomašica, Bosnia and Herzegovina.

FORENSIC SCIENCE

Bringing out the dead

Alison Abbott reviews the story of how a DNA forensics team cracked a grisly puzzle.

During nine sweltering days in July 1995, Bosnian Serb soldiers slaughtered about 7,000 Muslim men and boys from Srebrenica in Bosnia. They took them to several different locations and shot them, or blew them up with hand grenades. They then scooped up the bodies with bulldozers and heavy earth-moving equipment, and dumped them into mass graves.

It was the single most inhuman massacre of the Bosnian war, which erupted after the break-up of Yugoslavia and lasted from 1992 to 1995, leaving some 100,000 dead. With the war's end in sight, the Serbian army had to worry about hiding the evidence. In the late summer, they brought out the bulldozers again, roughly dug up the decaying bodies, threw them into dumper trucks and distributed them between 30 or so more remote burial sites. After the war shuddered to a halt in the autumn, these hastily disguised sites, with their cargoes of disconnected bones, were discovered. Christian Jennings's

Bosnia's Million Bones tells the story of how innovative DNA forensic science solved the grisly conundrum of identifying each bone so that grieving families might find some closure.

This is an important book: it illustrates the unspeakable horrors of a complex war whose causes have always been hard for outsiders to comprehend. The author, a British journalist, has the advantage of on-the-ground knowledge of the war and of the International Commission on Missing Persons (ICMP), an organization created in Sarajevo in 1996 that has a central role in the story. In 2000, the ICMP launched the world's first systematic attempt to apply DNA-identification techniques to large numbers of people. Its labs have since been used to help to identify individuals in other large groups killed in natural disasters, accidents and wars — including the 2013 terrorist attack on Nairobi's Westgate shopping centre, in which dozens of victims were mangled beyond conventional recognition.

As Jennings shows, the organization's first job was a masterwork from hell that involved

Bosnia's Million Bones: Solving the World's Greatest Forensic Puzzle
CHRISTIAN JENNINGS
Palgrave Macmillan: 2013.

locating, storing, preparing and analysing the million or more bones. It was in large part possible because during those fateful days in July 1995, aerial reconnaissance missions by the United States and the North Atlantic Treaty Organization had picked up images of large groups of men on open ground near Srebrenica. Subsequent images showed that the men had disappeared and large areas of disturbed earth had appeared. Over the following weeks, as the bodies were relocated, images showed more stretches where the soil was newly disturbed.

In 1997 and 1998, a team of archaeologists and forensic experts — put together by the Netherlands-based United Nations International Criminal Tribunal for the ▶

► former Yugoslavia — began excavating the burial sites. They pieced together some evidence of when and how the mass killings had taken place from clues such as the bodies' states of decay, the times and dates on their self-winding watches, and the characteristic patterns of damage caused to skulls by bullets. Analysis of the colours and textures of soils pointed to where some of the bones had first been dumped. For example, chips of glass indicated burial near a glass factory in the area.

The task of identifying the bones was exquisitely difficult. The bulldozers had broken up the bodies, and the pieces had been mixed up in the dumper trucks transporting them to new burial sites. DNA analysis of each bone was the only possible method of conclusive identification, so the ICMP set up its lab.

At first, this remarkable operation ran on a shoestring. Members invented cheap alternatives for equipment, such as adapting a chicken rotisserie from the local market to stir DNA solutions. All of these staff (many of them "massively adaptable" graduates, Jennings writes) were locals, who could easily communicate with the traumatized relatives of the missing. This helped them to collect the blood samples for the DNA analysis needed for comparison with DNA from the bones.

Each staff member was trained in a specific aspect of this analysis, which was then carried out in modular fashion. The remains were first prepared for DNA extraction, then ground into powder in the Republic of Srpska, now an independent Serbian enclave within Bosnia. Next, the powder was transferred to Sarajevo for DNA extraction. Through that analysis, more than 80% of the remains were returned to their families for burial.

> **"More than 80% of the remains were returned to their families for burial."**

That story needed to be told. But *Bosnia's Million Bones* is a confusing read. It weaves in other, undoubtedly important, stories — such as the manhunt for the war criminals responsible for the massacres — and diverts frequently into issues involving unrelated wars. Its structure is undisciplined, muddling timelines and sometimes even basic numbers (such as the number of victims identified by a particular date). But those who make it through will emerge shaken, and educated. ∎

Alison Abbott *is* Nature's *senior European correspondent.*
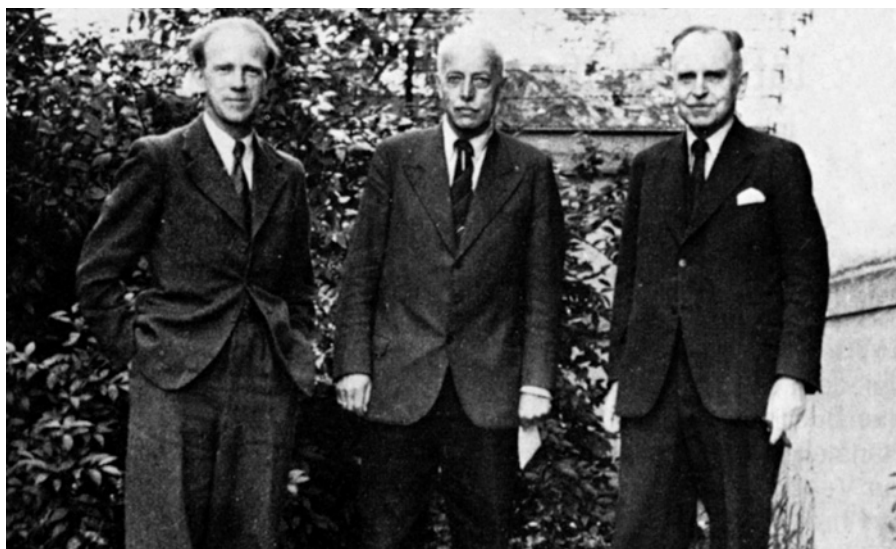
Left to right: **Werner Heisenberg, Max von Laue and Otto Hahn in Göttingen, Germany, in 1946.**

PHYSICS

Overhearing Heisenberg

Ann Finkbeiner ponders a script inspired by the 1945 internment of eminent German physicists in England.

By July 1945, the Allies and Germans had spent years racing each other to build an atomic bomb. The German physicists were certain of their technological superiority, but had not even taken the first step — building a working reactor. The Manhattan Project scientists, who had panicked that the Germans would build this evil thing first, had made four bombs. But that July, neither side knew for certain how close the other had come. So, just after the Nazi surrender, the Allies captured ten German nuclear scientists — including Werner Heisenberg, Otto Hahn, Max von Laue, Kurt Diebner and Carl Friedrich von Weizsäcker — sequestered them in Farm Hall, a country house in deepest Cambridgeshire, UK, and bugged their rooms.

Transcripts of the taped conversations were declassified and published almost 50 years later in *Operation Epsilon* (University of California Press, 1993) and annotated in physicist Jeremy Bernstein's *Hitler's Uranium Club* (AIP Press, 1995). But they begged to be a play. Now David Cassidy, historian of physics at Hofstra University in New York, has written a one-act script called *Farm Hall*. Whereas a recent produced play by Alan Brody (also called *Operation Epsilon*) focused on the scientists' morals in trying to build a bomb for Hitler, Cassidy looks at the scientists' accounts of their failure to do so.

Both playwrights had to choose, from the mess of reality, one central tension. I thought that the tension might be how close

Farm Hall
DAVID CASSIDY
Staged reading in Santa Fe, New Mexico: May 2014.

the Germans came to building the bomb. Bernstein thought the tension was the German scientists' construction of a version of reality in which they had refused to build the bomb for Hitler on principle. Cassidy, however, focuses on their realization of their technological inferiority — on how they rationalized what he calls their own "fall into failure".

Cassidy quotes verbatim from the transcript, putting the stiffly translated German into American English. He narrows the cast down to five scientists, including Heisenberg, who led the German nuclear programme and won the 1932 Nobel Prize in Physics; Hahn, who co-discovered fission; and Diebner, an engineer. Their military minder at Farm Hall, Theodore Rittner, has arranged for the secret taping, translation and transcription of their conversations for British and US intelligence.

The scientists settle in and get comfortable. They talk.

They try to figure out why they're being held. To keep them out of the hands of the Russians? Because the Allies want to know what they know? They compliment themselves on being ahead of the Allies, who — they think — cannot build a reactor in which uranium can be collected into a near-critical mass and begin fission. They argue about why they never actually built the reactor: because Heisenberg insisted on using his design rather

than Diebner's more effectual one?

The scientists skirt around the moral issue of building an atomic bomb for the Reich. Heisenberg and the others agree that they did what was necessary to protect the future of German science. Hahn, who never worked on the bomb, says that he loves Germany but is glad that her criminal leaders lost the war. Diebner says that he joined the Nazi party because he needed work.

On the night of 6 August, they listen to the BBC's announcement that the United States has dropped the atomic bomb on Hiroshima. Stunned, they try to figure out how the Allies managed it. Heisenberg calculates that by using 100,000 mass spectrometers, one could separate out enough of the fissile but rare isotope of uranium for a bomb — about a tonne. Hahn is confused: aren't Heisenberg's calculations out by a factor of ten? (They are.)

The next day, they read the British newspapers, which brag that the Allies won the atomic race. They are outraged, having thought they were so far ahead that racing was irrelevant. They disagree about whether they were even trying to build a bomb or, as Heisenberg begins to insist, just a reactor. Everybody agrees that the German government kept them too short of funds for success. They write an official memorandum explaining that their efforts were directed towards building a power-producing reactor and that working on a bomb had not been feasible. About five months later, they go home — Heisenberg to the directorship of the Kaiser Wilhelm Institute for Physics in Berlin, and the others also to worthy and interesting jobs. As Cassidy says, they fall from the heights of their arrogance, but not far.

Cassidy's script has had two readings; others are planned, and a Spanish production in Santiago, Chile, is in preparation. Cassidy is expanding his play to two acts. "I don't think I could have picked a more difficult subject for my first play," he says. The difficulty lies in the multiplicity of historical realities that he must cram into one plot that is driven, in effect, by one tension.

The transcript itself holds many tensions: between aristocratic theorists and lower-caste engineers; between those who joined the Nazis and those who just worked for them; between arrogance and wilful blindness; between Heisenberg's great scientific stature and his failure to help a Jewish colleague's family, or indeed his own. Cassidy has Rittner, at the play's end, collapse all the tensions: people who are great in one area, Rittner says, are expected to be — and expect themselves to be — great in all. But in both art and life, they fall. ∎

Ann Finkbeiner *is a freelance science writer in Baltimore, Maryland, and author of* The Jasons.
*e-mail: anniekf@gmail.com*

Books in brief

Shaping Humanity: How Science, Art, and Imagination Help Us Understand Our Origins
*John Gurche* YALE UNIVERSITY PRESS *(2013)*
Palaeoartist John Gurche crafts hyperrealistic sculptures of extinct hominins, built up from casts or three-dimensional models of their skeletons. To bring these individuals from deep time to 'life', Gurche fuses his knowledge of comparative anatomy with forensic science and informed guesses about expressions and poses. His coffee-table gem showcases and contextualizes 15 of these finely judged creations, representing a span of 6 million years and ranging from *Sahelanthropus tchadensis* to the 'Hobbit' *Homo floresiensis*.
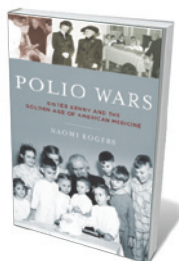
Polio Wars: Sister Kenny and the Golden Age of American Medicine
Naomi Rogers OXFORD UNIVERSITY PRESS *(2013)*
Before the Salk vaccine was licensed in 1955, polio epidemics swept the United States. Naomi Rogers traces them through the story of Australian-born 'bush nurse' Elizabeth Kenny, who eschewed splinting in favour of early muscle manipulation. Her star rose, but her methods stirred controversy and she was forgotten with the vaccine's advent. Kenny's principal legacy, Rogers speculates, might be her idea — unacknowledged in the evolution of polio science — that the disease was systemic rather than neurotropic.
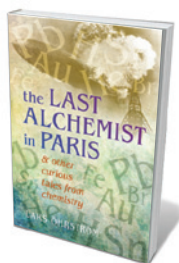
The Last Alchemist in Paris: And Other Curious Tales From Chemistry
*Lars Öhrström* OXFORD UNIVERSITY PRESS *(2013)*
History offers a painless way to grasp the periodic table's 114 confirmed elements, notes chemist Lars Öhrström. So, for instance, we visit Cumbria in northern England, once an "information technology hub" that supplied the graphite used in pencils. And we follow the Swedish playwright August Strindberg as, gripped by psychosis, he set up an alchemical lab in Paris — leading Öhrström to ponder lithium carbonate (used to treat bipolar disorder), as well as gold. There is much more in this charming mishmash of a primer.

Fritz Kahn
*Uta von Debschitz and Thilo von Debschitz* TASCHEN *(2013)*
The 1926 *Man as Industrial Palace* is only the most iconic of the images unleashed by infographics pioneer Fritz Kahn. A modernist genius, Kahn's illustrations were endlessly inventive, often darkly comic and occasionally macabre. His 1924 drawing *Travel Experiences of a Wandering Cell: In the Valley of a Flesh Wound*, for example, beautifully elucidates the living landscape of blood, nerves and tissue. In this biography in English, German and French that features 350 of his works, Uta and Thilo von Debschitz pay homage to the half-forgotten artist on the 125th anniversary of his birth.
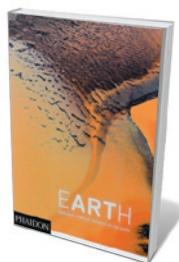
Earthart: Colours of the Earth
Bernhard Edmaier and Angelika Jung-Hüttl PHAIDON *(2013)*
Distance lends enchantment to Earth's particoloured, pitted surface, as revealed by this photofest by two geologists, writer Angelika Jung-Hüttl and photographer Bernhard Edmaier. Terrestrial meanders, fractals and waves echo biological forms, and vivid hues remind the reader how earthly muds and minerals yield pigments from yellow ochre to ultramarine. A chance to enter an alternative vision of our planet, from the smoked-glass icebergs of East Greenland to the stupendous lion-coloured reaches of the Chilean Andes. **Barbara Kiser**
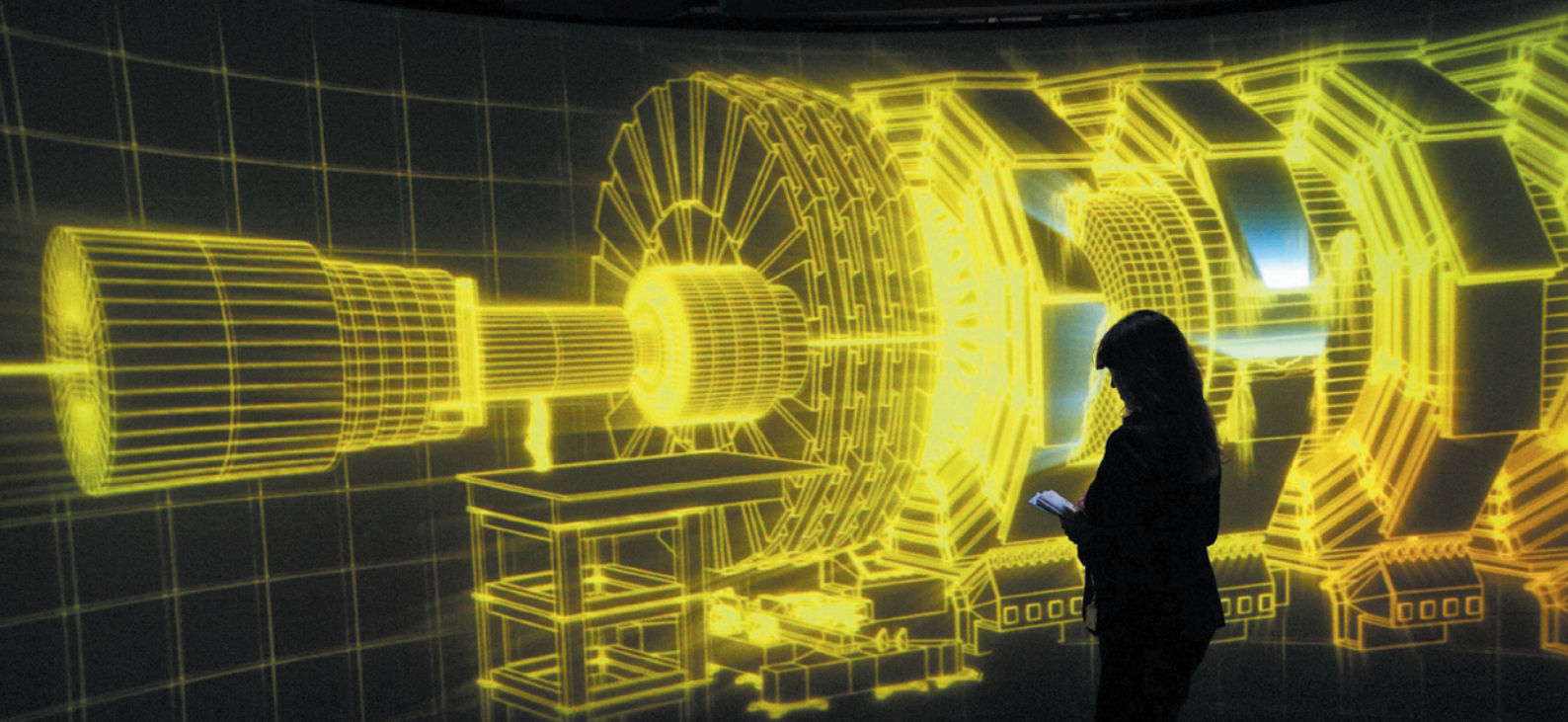
A video projection at the London Science Museum's *Collider* exhibition takes visitors up close to CERN's particle accelerator.

PARTICLE PHYSICS

Smashing spectacle

Zeeya Merali weighs up a simulated tour of the Large Hadron Collider.

It is the world's biggest, most expensive science experiment, smashing protons together at the highest energies reached on Earth. The Large Hadron Collider (LHC) — the particle accelerator at CERN, Europe's high-energy physics laboratory near Geneva in Switzerland — feeds on superlatives. So curators attempting to recreate the experience of walking through its underground tunnels, which occupy a space 27 kilometres in diameter, faced a seemingly impossible task. But through a mixture of animation, audio and video, and a handful of objects, the London Science Museum's £1-million (US$1.6-million) *Collider* exhibition just about pulls it off.

Targeting the over-16s, and consciously aiming to encourage budding physicists, the exhibition avoids the museum's trademark interactive games in favour of immersing visitors in the sights, sounds and culture of CERN. The focus is on the human motivation behind the machines, rather than on the theoretical physics or the accelerator's technological achievements.

The tour begins in a replica of CERN's auditorium. A short film reimagines the July 2012 announcement of the discovery of the Higgs boson — the final missing piece in the standard model of particle physics — that explains how matter acquires mass. Despite creditable performances, the actors could not quite deliver the enthusiasm that the LHC staff themselves showed in news reports; it would also have been welcome to

Collider Exhibition
SCIENCE MUSEUM,
LONDON
Until 6 May 2014.

see a wider diversity of faces in the film.

Next, visitors — like the LHC's particle beams — must move along a tightly prescribed curved path through the exhibition space. It was, of course, impossible to create a facsimile of the awe-inspiring underground caverns, lined with vast, intricate machinery. Nevertheless, a canny soundscape generates an authentic atmosphere, layering snippets of conversations between workers against the background hum of the detector. Loving attention to the low-tech detail that typifies much of CERN, including outdated computer towers and bicycles used to traverse the circuit, also lend credibility to the exhibition. Dull grey hallways lead to the office of a physicist who clearly eats and sleeps at her desk, conjuring the daily grind of science in action.

Along the route, traditional glass display cases contain small items of apparatus, such as hydrogen canisters and sections of magnets. These are brought to life by engaging life-size video footage of LHC researchers mixing technical explanations with anecdotes about working in the world's largest scientific collaboration. Charming hand-drawn diagrams, humorous cartoons and simple animations drip-feed background physics.

A highlight is the 270-degree wrap-around projection of the heart of the collider. Graphics zoom in to proton collisions and then back

out to depict experimental data as it is transmitted around the globe, helping the viewer to visualize the collider's extremes of scale.

The most detailed physics explanations are saved for the end. Here, we learn more about the search for the Higgs boson (which involved two years of collisions and some 10,000 people), such as how its presence was inferred through data analysis. In the final room, a delightful series of animations dance atop a white desk, vividly tracing out some of the biggest mysteries — such as the nature of dark matter, which the collider is now harnessed to crack. Enhanced by voice-overs from physicists, including CERN project leader Lyn Evans, this is a simple yet evocative way to convey complex theories. Both these voice-overs and the use of actual researchers in the earlier video clips hint at a lost opportunity: the opening film might have benefited from the inclusion of real physicists, rather than actors, talking about their excitement over the Higgs announcement.

Collider is not the place to learn in-depth physics. Nor should it be. Instead, it succeeds in showcasing a monumental scientific endeavour from a human perspective — and leaves visitors hungry to find out more. ∎

Zeeya Merali *is a freelance science writer based in London, and editor for the Foundational Questions Institute in Georgia, USA.*
e-mail: merali@fqxi.org

Correspondence

Mitigate damage risk from bush fires

As climate models project a severely increased fire risk in southeastern Australia (H. G. Clarke *et al. Int. J. Wildland Fire* **20,** 550–562; 2011), we urgently need to put a stop to social processes that are amplifying our risk from bush fires (R. P. Crompton *et al. Wea. Climate Soc.* **2,** 300–310; 2010). More thoughtful land use and planning could curb the destruction they cause in and around cities.

Interviews with residents following the recent fires in Sydney echoed those after the devastating Black Saturday bush fires in the state of Victoria in 2009 (J. Whittaker *et al. Int. J. Wildland Fire* **22,** 841–849; 2013). These residents likewise complained that government restrictions prevented them from clearing trees on their land as a fire break and that suburbia continues to expand into untamed bush land without due preparedness and protection.

The recent blazes in Sydney occurred earlier in the fire season than usual, prompting speculation about a possible link with climate change.

Katharine Haynes, Deanne Bird, John McAneney *Risk Frontiers, Macquarie University, Sydney, Australia.*
haynes.katharine@gmail.com

Russian universities need change of tack

Of the factors contributing to Russia's poor scientific performance (see, for example, A. Gorobets *Nature* **503,** 39; 2013), the plight of the country's university professors stands out.

These professors are poorly paid; most are forced to supplement their earnings with other employment. Their typical annual teaching load of up to 1,000 hours is three to five times higher than that in universities elsewhere, leaving them little time for research or writing up their work. On top of this, professors are now under pressure from Russia's Ministry of Education and Science to publish two or three papers a year in international journals to help their universities to reach the top 200 in global rankings.

By contrast, the world's leading universities do not expect their professors to teach for more than 200–400 hours a year, allowing them enough time for research. They also protect the intellectual property that results from this research, which contributes substantially to university budgets — amounting in some US universities to several billions of dollars a year (roughly the combined research budget for some provinces and republics in the Russian Federation).

A transfer to universities of institutions currently owned by the beleaguered Russian Academy of Sciences (see *Nature* http://doi.org/p6d) might offer a solution. It would empower university research, help to dissolve Russia's rigid and detrimental education–research divide, and ultimately boost the rankings of Russian universities.

Renad Zhdanov *Kazan Federal University, Russia, and Sholokhov Moscow State University for the Humanities, Moscow, Russia.*
zrenad@gmail.com

Definition of maths genius is elusive

'Project Einstein' aims to identify the genotypes of prominent mathematicians (*Nature* **502,** 602–603; 2013), but it needs to be underpinned by an accurate definition of the mathematical-genius phenotype.

Basic mathematical competence is judged according to numeracy and arithmetical skills. Advanced ability is less easily delineated; genius-level mathematical ability is even harder to define. Advanced mathematics encompasses diverse elements such as sophisticated abstract thought, statistical know-how, raw computation, geometric awareness, imagination, lateral thinking, logic and philosophy. Moreover, proficiency in all of these areas has yet to be properly quantified in terms of inherent versus learned ability.

So the genetic heterogeneity studied through gene sequencing is unlikely to arrive at a 'mathematical-genius genome'. Such studies may, however, shed light on frequently associated neurodevelopmental conditions (see, for example, S. Baron-Cohen *et al. J. Autism Dev. Disord.* **31,** 5–17; 2001).

Hutan Ashrafian *Imperial College London, UK.*
h.ashrafian@imperial.ac.uk

Greece's high CT scanning record

You remark on the high number of computed tomography (CT) scans used in Greece, but only risk–benefit analysis and cost-effectiveness studies will indicate whether this is a good or a bad thing (*Nature* **502,** S82–S83; 2013). Your claim that the country has no official guidelines governing the use of CT scans is not correct.

CT scanners in Greece have been regularly monitored under strict guidelines since 2001. And diagnostic and therapeutic protocols in radiology were implemented in 2011.

Peculiarities in Greece's health system contribute to the high level of CT scanner use. The number of doctors per capita is almost double the average for countries in the Organisation for Economic Co-operation and Development, which may result in overprescription of diagnostic procedures. This could be due in part to the relatively low cost of CT scans in Greece (US$88 for a chest scan, for example, compared with $332 in the United States). A new electronic referral system has now been set up that should discourage overprescribing.

Ioannis Seimenis *University of Thrace, Greece.*
Stelios Argentos, Stathis Efstathopoulos *University of Athens, Greece.*
stathise@med.uoa.gr

TB vaccine failure was predictable

Your report on tuberculosis (TB) vaccines perpetuates a flawed but widely held view. In fact, the lack of efficacy of the MVA85A vaccine in a recent human clinical trial was no surprise (*Nature* **502,** S8–S9; 2013).

The trial followed an immunization regime previously used in four animal models in the past ten years. Careful examination of those results reveals that MVA85A offered no statistically significant increase in protection over the BCG (Bacillus Calmette–Guérin) vaccine alone in mice, guinea pigs, cows and non-human primates (see, for example, F. A. W. Verreck *et al. PLoS ONE* **4,** e5264; 2009; and S. A. Sharpe *et al. Clin. Vaccine Immunol.* **17,** 1170–1182; 2010).

The only exception is a mouse study (N. P. Goonetilleke *et al. J. Immunol.* **171,** 1602–1609; 2003), which is not comparable because a different immunization route was used (see C. N. Horvath and Z. Xing *Adv. Exp. Med. Biol.* **783,** 267–278; 2013).

In aggregate, therefore, the preclinical animal data for MVA85A predicted the outcome of the reported clinical trial. It remains to be seen how successfully animal models will predict the efficacy of other TB vaccine candidates.

Peter Beverley *University of Oxford, UK.*
peter.beverley@ndm.ox.ac.uk

George Herbig
(1920–2013)

Astronomer who pioneered studies of young stars.

George Herbig's research, which spanned more than 70 years, built the foundation on which rests our present-day understanding of the birth of stars and of the properties of young stars. He had an uncanny ability to identify astronomical objects and research topics that would become key elements in the study of early stellar evolution.

Herbig, who died on 12 October, was an only child born in modest circumstances in Wheeling, West Virginia. His father, a tailor, had settled there after emigrating from Germany. Sometime after his father's early death, Herbig moved to Los Angeles, California, where as a teenager he built his first telescope. The nearby Mount Wilson Observatory, housing what was then the world's largest telescope, with a 2.5-metre mirror, fostered his growing interest in astronomy.

Through joining the Los Angeles Astronomical Society as a young man, Herbig met many of the great astronomers of the time, and had the opportunity to attend observations at Mount Wilson. He later spoke of the awe he had experienced when looking, using the spectrograph slit of the 2.5-metre telescope, at the giant star Mira — a luminous red spot, seemingly boiling as a result of its light passing through Earth's turbulent atmosphere. At the tender age of 20 he published his first brief scientific results, on the diameter of stars.

From observations spanning from the late 1930s to the early 1940s, Herbig's mentor, Alfred Joy, had discovered a peculiar class of variable stars named after the prototype star T Tauri. These objects are often associated with dark interstellar clouds, and it was initially speculated that their characteristic variable brightness could be attributed to the stars passing through the gas and dust of the interstellar medium.

T Tauri stars became the topic of Herbig's 1948 PhD thesis, *A Study of Variable Stars in Nebulosity*. His work supported the growing consensus that these stars are very young — with their luminosity arising not from nuclear burning, but from the release of energy as the stars contract under gravity. Following decades-long systematic studies of T Tauri stars, Herbig synthesized, in 1962, all that was known at the time about the class in a now-famous paper, 'The Properties and Problems of T Tauri Stars and Related Objects', which has become the foundation for the modern study of these young stars (G. H. Herbig *Adv. Astr. Astrophys.* **1**, 47–103; 1962).

As part of his investigation of T Tauri stars, Herbig studied a region of dark clouds in the Orion constellation, in which he noticed small nebulous objects with peculiar spectra. This class is now known as Herbig–Haro objects, after Herbig and astronomer Guillermo Haro, who had independently discovered them. Over several decades of study, Herbig and his collaborators established that Herbig–Haro objects move with supersonic velocities away from newborn stars, and are thus the signposts of recent star-formation events.

T Tauri stars are low-mass stars that eventually become similar to or smaller than the Sun. Herbig recognized that counterparts of these young stars, with masses several times that of the Sun, ought to exist as well. After exhaustive studies, he published in 1960 a landmark paper describing the discovery and characterization of the more-massive stars, now known as Herbig Ae and Be stars. Observations with telescopes, both ground-based and spaceborne, have revealed that disks of debris can surround these stars, and that in some cases these disks harbour newly formed planets and cometary bodies. As sites of planetary genesis, these Herbig stars have in recent years become the subject of intense study.

Herbig was fascinated by stars that are oddballs, recognizing that because stars live so much longer than humans, important evolutionary stages — if brief enough — may be seen only very rarely. In 1936, a faint variable star, FU Orionis, brightened 100-fold within six months, and has barely declined in luminosity since. Herbig studied this star and similar cases, and realized that such events represent important episodes in the early lives of some stars. Unafraid to take a stand against the prevailing wisdom, Herbig maintained that these 'FUor' events represent rapidly rotating young stars near the point of break-up. Most in the community believe that such events are the result of heating in a surrounding disk, which makes the disk self-luminous. But there are now signs that a hybrid model combining both these aspects might explain what is actually happening.

At an age when most people retire, Herbig embarked with his students on a series of observational studies of clusters of very young stars — groups of many hundreds or thousands of stars born together. He espoused the idea that star formation in clusters proceeds over several millions of years, with most low-mass stars forming first, until the birth of very energetic massive stars suddenly destroys the clouds of gas and dust from which stars are born and brings further star formation to a rapid halt.

Modest, mild-mannered and softly spoken, George exuded a quiet authority. He was an independent and private man, usually observing alone, and commonly processing and analysing the data himself. During his long career he saw major transformations in instrumentation and techniques — such as from photographic plates guided by eye to charge-coupled device cameras on telescopes controlled by computers.

We would sometimes joke that we had misspent our lives; we could have stayed at the pub while all the wonderful new hardware and software was being developed, and then have accomplished in a few years what had taken a lifetime. Of course, it is only in hindsight that there seems to be a shortcut in the winding path to knowledge and discovery. ■

Bo Reipurth *is an astronomer at the University of Hawaii, and worked closely with George Herbig in his later years. e-mail: reipurth@ifa.hawaii.edu*

KAREN TERAMURA/UNIV. HAWAII INST. FOR ASTRONOMY

Does quadrupole stability imply LLSVP fixity?

ARISING FROM C. P. Conrad, B. Steinberger & T. H. Torsvik *Nature* **498**, 479–482 (2013)

The African and Pacific large low-shear-velocity provinces (LLSVPs) at present dominate the structure of the Earth's lowermost mantle, but there is considerable debate as to whether these structures have remained fixed throughout geologic time or whether they have shifted in response to the changing configurations of mantle downwellings associated with zones of surface tectonic plate convergence. In a recent Letter, Conrad *et al.*[1] performed a multipole expansion of the Earth's plate motions from 250 million years (Myr) ago to the present and used the relatively stationary positions of quadrupole divergence to argue that the two LLSVPs have remained stationary at least for the past 250 Myr and further speculated that the two LLSVPs formed "stable anchors" in the more distant geologic past. Here we show that the quadrupole divergence of plate motions is not a representative diagnostic for overall plate divergence patterns, owing to cancellation effects in the multipole expansion. Hence, the conclusion by Conrad *et al.*[1] that the presence of stationary quadrupole divergence implies fixity of the LLSVPs is not well founded. There is a Reply to this Brief Communication Arising by Conrad, C. P., Steinberger, B. & Torsvik, T. H. *Nature* **503,** doi:10.1038/nature12793 (2013).

Conrad *et al.*[1] define "net characteristics" of plate tectonics on the Earth based on the dipole and quadrupole contributions to the plate motions. These net characteristics are very similar to the spherical harmonic representation of the poloidal component of the plate motions, which represents convergent and divergent motion on the sphere[2,3]. It can be shown that the pure dipole in ref. 1 is identical within a multiplicative constant to the $(1, 0)$ spherical harmonic of the divergence field when the dipole axis is aligned with the axis of rotational symmetry in the spherical coordinate system, and that the pure quadrupole is similar to the $(2, 2)$ spherical harmonic contribution. In Figure 1 of this Comment, we show the degree-1 and degree-2 contributions to the plate motions at 200 Myr ago (left column) from ref. 4, similar to those in ref. 1 and at 300 Myr ago (right column) from ref. 5. The locations of degree-1 and degree-2 spherical harmonic extrema (circles and diamonds in Fig. 1) are very similar to the dipole and quadrupole orientations in figure 3 of ref. 1.

Divergent plate motion in the African hemisphere (that is, within Pangaea) may have started around 290 Myr ago with the Neo-Tethys seafloor spreading[6], but the divergence in the African hemisphere was much weaker than in the Pacific hemisphere at 200 Myr ago (bottom row of Fig. 1). However, degree-2 divergence has equal amplitude in the two hemispheres (second row of Fig. 1). When we examine spherical harmonic degrees up to and including 40 (bottom row of Fig. 1), we can see that the apparent degree-2 divergence in the African hemisphere is largely cancelled by other modes (including degree-1). This suggests that in general the degree-2 divergence alone is not a good proxy for the long-wavelength structure of plate motions.

In fact, the degree-2 divergence field for a proxy plate motion model at 300 Myr ago (right column in Fig. 1) is similar in both amplitude and orientation to that at 200 Myr ago, despite the complete absence of divergence in the African hemisphere[5,6]. At 300 Myr ago, we assume[5] that seafloor spreading in the Panthalassic hemisphere was accommodated by circum-Pangaea subduction. The amplitude of degree-2 motion at 300 Myr ago (4.45 cm per year) is greater than at 200 Myr ago (3.47 cm per year). However, at 300 Myr ago, the degree-1 convergence maximum is closely aligned with the degree-2 divergence maximum located within Pangaea, resulting in cancellation. Because we are concerned chiefly with the long-wavelength characteristics of the plate motions, only the presence of spreading (not the precise details of plate motions) in the Panthalassic hemisphere is important.
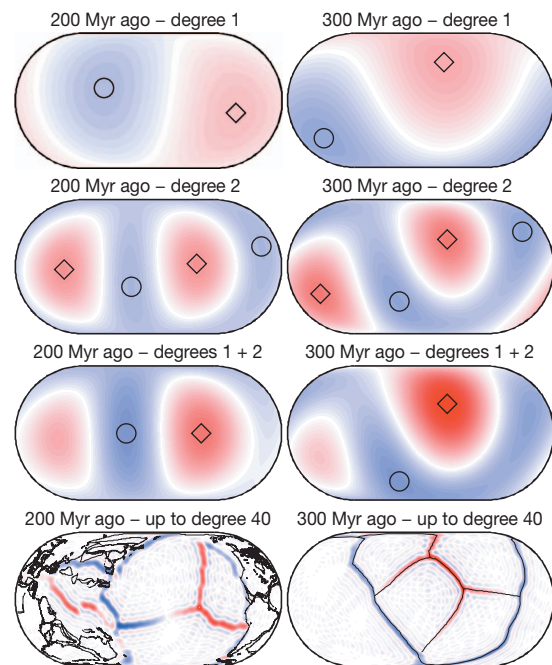

Figure 1 | Spherical harmonic contributions to the poloidal velocity field. Top row, for spherical harmonic degree 1; second row, for degree 2 only; third row, for degrees 1 and 2; and bottom row, for all degrees up to 40. Circles indicate convergence maxima; diamonds indicate divergence maxima.

Although we agree with ref. 1 that plate motions provide constraints on changes in mantle buoyancy structure, the plate divergence in the African hemisphere has changed from strongly positive since the break-up of Pangaea to weakly positive at around 200–250 Myr ago, becoming neutral at around 300 Myr ago, and was most probably negative before the formation of Pangaea at 330 Myr ago (that is, reflecting convergence associated with the assembly of Pangaea). This clearly suggests that mantle buoyancy forces in the African hemisphere have changed significantly during the last supercontinent cycle. In laboratory and numerical studies, cold downwelling material that reaches the core–mantle boundary tends to push aside compositionally dense material, a process that is thought to be analogous to the interaction of LLSVPs with subducted oceanic crust[5,7–9]. Hence, changes in mantle buoyancy structure are expected to change the LLSVP arrangement, and numerical simulations driven by velocity boundary conditions successfully reproduce the present-day arrangement of the LLSVPs[5,7,10] while allowing the LLSVPs to shift in response to changes in downwelling structure. Robust observations are needed to test the time evolution of the LLSVP structures, and the quadrupole divergence component alone is not a sufficiently robust indicator of the past mantle flow field to assess the long-term fixity of the LLSVPs.

METHODS

We calculated the surface divergence and radial vorticity of the velocity field defined by plate reconstructions for 200 Myr ago[4] and defined by a proxy plate reconstruction prior for 300 Myr ago[5] using CitcomS[11] and performed spherical harmonic expansion of the resulting scalar fields representing poloidal and toroidal motion, respectively[2,3]. The amplitudes of individual spherical harmonic degrees reported in the text are calculated using the normalizations introduced in ref. 3. (This

BRIEF COMMUNICATIONS ARISING

Maxwell L. Rudolph[1] & Shijie Zhong[1]
[1]Department of Physics, University of Colorado, Boulder, Colorado 80309, USA.
e-mail: maxwell.rudolph@colorado.edu

1. Conrad, C. P., Steinberger, B. & Torsvik, T. H. Stability of active mantle upwelling revealed by net characteristics of plate tectonics. *Nature* **498**, 479–482 (2013).
2. Hager, B. H. & O'Connell, R. J. Subduction zone dip angles and flow driven by plate motion. *Tectonophysics* **50**, 111–133 (1978).
3. Lithgow-Bertelloni, C., Richards, M. A., Ricard, Y., O'Connell, R. J. & Engebretson, D. C. Toroidal-poloidal partitioning of plate motions since 120 MA. *Geophys. Res. Lett.* **20**, 375–378 (1993).
4. Seton, M. *et al.* Global continental and ocean basin reconstructions since 200Ma. *Earth Sci. Rev.* **113**, 212–270 (2012).
5. Zhang, N., Zhong, S., Leng, W. & Li, Z.-X. A model for the evolution of the Earth's mantle structure since the Early Paleozoic. *J. Geophys. Res.* **115** (B6), B06401 (2010).
6. Lawver, L. A., Coffin, M. F., Gahagan, L. M., Campbell, D. A. & Royer, J. Y. PLATES 2004 atlas of plate reconstructions (750 Ma to present day). PLATES Project Report 207 (Institute for Geophysics, Univ. of Texas, 2004).
7. Bower, D. J., Gurnis, M. & Seton, M. Lower mantle structure from paleogeographically constrained dynamic Earth models. *Geochem. Geophys. Geosyst.* **14**, 44–63 (2013).
8. Jellinek, A. M. & Manga, M. Links between long-lived hot spots, mantle plumes, D″, and plate tectonics. *Rev. Geophys.* **42**, RG3002 (2004).
9. Tan, E., Leng, W., Zhong, S. & Gurnis, M. On the location of plumes and lateral movement of thermochemical structures with high bulk modulus in the 3-D compressible mantle. *Geochem. Geophys. Geosyst.* **12**, Q07005 (2011).
10. McNamara, A. K. & Zhong, S. Thermochemical structures beneath Africa and the Pacific Ocean. *Nature* **437**, 1136–1139 (2005).
11. Zhong, S., McNamara, A., Tan, E., Moresi, L. & Gurnis, M. A benchmark study on mantle convection in a 3-D spherical shell using Citcom S. *Geochem. Geophys. Geosyst.* **9**, Q10017 (2008).

Conrad *et al.* reply

REPLYING TO M. L. Rudolph & S. Zhong *Nature* **503**, http://dx.doi.org/10.1038/nature12792 (2013)

We thank Rudolph and Zhong for their Comment[1], which allows us to highlight important aspects of our original Letter[2]. In particular, they have provided an example of plate motions at 300 million years (Myr) ago (see right column of figure 1 of ref. 1) in which the plate tectonic quadrupole is not representative of plate tectonic divergence patterns (that is, there is no divergence in the middle of their supercontinent, despite a divergent quadrupole there). However, our study[2] does not claim that there should be a correspondence between quadrupole locations and the specific locations of plate tectonic divergence—instead we argue that plate tectonic dipole and quadrupole locations are representative of underlying mantle flow only.

This 300-Myr example actually demonstrates the utility of our method. To see this, consider mantle flow beneath a supercontinent covering one-third of the globe: mantle upwelling is expected beneath the opposing ocean's spreading ridges, but mantle downwelling occurs neither opposite to this upwelling (as for dipole flow) nor in bands 90° away (as for quadrupole flow), but instead associates with subduction occurring between these two locations on the supercontinent's periphery. Return flow from this downwelling should drive upwellings beneath both the supercontinent and the oceanic plates. Indeed, upwelling is expected beneath a supercontinent that will soon disperse[3,4].

Thus, we expect strong upwelling beneath the oceanic plates and weaker upwelling beneath the supercontinent; such a flow field is described by a combination of dipole and quadrupole flow fields that partially cancel beneath the supercontinent. This pattern is exactly predicted by the net characteristics (or spherical harmonics) of surface plate motions: the 300-Myr analysis of ref. 1 shows weak divergence within the supercontinent, indicating underlying upwelling. Thus, the lifetime of the two antipodal upwellings in the mantle may extend beyond the 250 Myr that we demonstrated in our original Letter[2]. More importantly, this analysis[1] demonstrates the importance of using only the longest-wavelength components of plate motions to visualize the underlying mantle flow patterns. By including shorter-wavelength spherical harmonic degrees, Rudolph and Zhong have incorporated the influence of regional and local tectonics into their interpretation[1]; doing this obscures the underlying mantle flow patterns that are only apparent at the longest wavelengths[2].

We agree with the Comment[1] that quadrupole stability alone does not prove long-term stability of the LLSVP regions, and that additional constraints from "robust observations" are necessary. Indeed, the locations of large igneous provinces and kimberlites have been shown to source from the margins of two antipodal LLSVPs[5], and would arise above a cold downwelling on the African side if the degree-1 interpretation of ref. 1 is correct. Furthermore, the 300-Myr plate motion example[1] is based on a study[6] that does not control for palaeolongitude or true polar wander, so it is unclear how surface features are related to LLSVP locations. Their portrayal of Pangaea as a stable coherent polygon additionally ignores much of the tectonic complexity of that supercontinent's evolution[7]. These problems illustrate the importance of using a carefully reconstructed model of past plate motions when attempting to use "net characteristics" to constrain LLSVP stability.

Clinton P. Conrad[1], Bernhard Steinberger[2,3] & Trond H. Torsvik[3,4,5]
[1]Department of Geology and Geophysics, SOEST, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA.
e-mail: clintc@hawaii.edu
[2]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.
[3]Centre for Earth Evolution and Dynamics (CEED), University of Oslo, 0316 Oslo, Norway.
[4]Geodynamics, Geological Survey of Norway, 7491 Trondheim, Norway.
[5]School of Geosciences, University of Witwatersrand, WITS 2050, Johannesburg, South Africa.

1. Rudolph, M. L. & Zhong, S. Does quadrupole stability imply LLSVP fixity? *Nature* **503**, http://dx.doi.org/10.1038/nature12792 (2013).
2. Conrad, C. P., Steinberger, B. & Torsvik, T. H. Stability of active mantle upwelling revealed by net characteristics of plate tectonics. *Nature* **498**, 479–482 (2013).
3. Gurnis, M. Large-scale mantle convection and the aggregation and dispersal of supercontinents. *Nature* **332**, 695–699 (1988).
4. Lowman, J. P. & Jarvis, G. T. Effects of mantle heat source distribution on supercontinent stability. *J. Geophys. Res.* **104**, 12733–12746 (1999).
5. Torsvik, T. H., Burke, K., Steinberger, B., Webb, S. C. & Ashwal, L. D. Diamonds sourced by plumes from the core–mantle boundary. *Nature* **466**, 352–355 (2010).
6. Zhang, N., Zhong, S., Leng, W. & Li, Z. X. A model for the evolution of the Earth's mantle structure since the Early Paleozoic. *J. Geophys. Res.* **115**, B06401 (2010).
7. Torsvik, T. H. *et al.* Phanerozoic polar wander, paleogeography and dynamics. *Earth Sci. Rev.* **114**, 325–368 (2012).

NEWS & VIEWS

A leak in the loop

The shifting nature of positive and negative feedbacks in a woodland region invaded by an exotic grass sheds light on the complexity of managing natural systems. SEE LETTER P.517

KATHARINE N. SUDING

Positive feedback typically refers to something most of us seek: a sign from others that we have done a good job. In systems theory, however, positive feedback does not necessarily have good connotations, and flattery may or may not produce it. Positive feedback in terms of a system — be it mechanical, economic, social or ecological — simply refers to a condition that is self-reinforcing, producing fast, amplifying change[1]. The sound produced by a microphone quickly grows if it is placed near a loudspeaker, for example, and a stampede starts with one panicked cow. On page 517 of this issue, Yelenik and D'Antonio[2] describe one such feedback mechanism, in which an invasive plant species changes the environment to its own benefit, increasing its abundance and furthering its own incursion.

Exotic plant species have been shown to alter several aspects of the ecosystem that they invade, including dynamics related to disturbance, hydrology and nutrient cycling[3]. When such effects promote or maintain the invader's dominance, a positive-feedback loop is formed. An example comes from sites in Hawaii where, in the 1960s, woodlands dominated by native *Metrosideros polymorpha* trees were invaded by the exotic grass *Melinis minutiflora*. Work at these sites in the 1990s was among the first to demonstrate enhanced invasion due to positive feedback or, in this case, due to two related feedbacks. First, the exotic grasses fuelled fires, which killed *Metrosideros* trees; this lack of trees led to the growth of more grasses, which fuelled more fires[4]. Second, the exotic grasses accelerated nitrogen-cycling rates, and more soil nitrogen increased the growth of the exotic grass[5]. Together, these feedbacks helped to convert *Metrosideros* woodlands to exotic grasslands across Hawaii[6].

One feature of positive feedback that is often overlooked is the fact that it cannot be sustained forever. Unchecked, a system trapped in a positive loop destroys itself; for example, a temperature-dependent chemical reaction produces heat and explodes. More commonly, positive feedback is checked by the development of negative feedback, leading to self-correction — when we get too hot, for
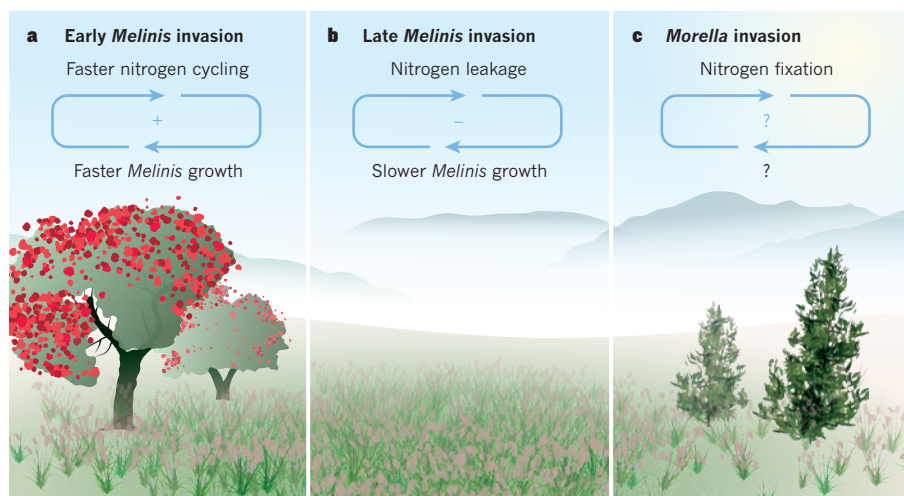

Figure 1 | A timeline of feedback. **a**, The invasion of the exotic grass *Melinis minutiflora* into *Metrosideros polymorpha* woodlands in Hawaii led to faster nitrogen cycling, which created a positive feedback that enhanced grass growth. **b**, Yelenik and D'Antonio[2] now show that leaking of nitrogen from invaded grasslands led to the development of a negative feedback, which slowed the grass invasion. **c**, Experiments performed by the researchers suggest that this negative feedback allowed the subsequent invasion of another exotic species, *Morella faya*, a nitrogen-fixing tree. Addition of nitrogen to the soil through fixation may allow a replay of the grass-invasion cycle, or an opportunity for restoration.

example, we sweat and we cool down.

In the first study of its kind, Yelenik and D'Antonio revisited the invaded sites in Hawaii to investigate whether the earlier positive feedback involving nitrogen had shifted over time. Studying several locations in the Hawaii Volcanoes National Park, they found that it had: the exotic grassland had developed a nitrogen leak, and the previously elevated nitrogen-cycling rates had dropped back to pre-invasion levels (Fig. 1).

As is typical of fast-growing grasses, *Melinis* produces a lot (often more than 2,000 grams per square metre) of relatively nitrogen-rich leaves. As winter approaches, these leaves die and break down in the soil. Leaf nitrogen then decomposes back to plant-available inorganic forms, feeding more grass growth. The leak probably occurred because winter rains can flush nitrogen out of the soil root zone before the plants start growing again. This mismatch between nitrogen release and uptake caused a negative-feedback loop to form: as progressively more nitrogen leaked from the system, the growth of the exotic grasses became more limited by nitrogen, and their invasion slowed.

The development of a negative feedback that slows an invasion might be viewed as good news for conservation. But the decline of an invader is only half the battle — recovery of the native species is also crucial. To assess this, Yelenik and D'Antonio planted a mix of native and exotic species and then simulated different phases of invasion, adding nitrogen to simulate the initial positive soil feedback and removing the invader to simulate its eventual decline. Two species stood out as benefiting the most from a low-nitrogen, invader-free environment: the native *Acacia koa* and the exotic *Morella faya*, both trees that have root bacteria able to fix atmospheric nitrogen.

Unfortunately for conservation, however, when the authors went to see which species were in fact colonizing the exotic grasslands following the decline of *Melinis*, they found only *Morella* trees (Fig. 1). This is worrisome, because it indicates that the positive feedback initiated by one invader had changed the organization of the system — perhaps tipping it into a new state[7] — and made way for another invader.

Yet possible restoration solutions exist. Yelenik and D'Antonio posit that the native *Acacia* trees should be able to establish themselves in the declining exotic grassland, but their heavy seeds just cannot get there. By contrast, the seeds of the exotic *Morella* are bird-dispersed, so this species easily wins the colonization race. But the race could be fixed by seeding *Acacia* into the exotic grassland, where it is likely to grow well. The question then arises of whether addition of nitrogen to the soil by the trees' fixation will reset the system and allow a replay of the grass invasion cycle. If so, the period of negative feedback could represent an opportunity to further reduce exotic-grass abundance to such a point that there is insufficient grass in the newly wooded regions to carry

fire, thereby minimizing the return of the initial positive feedback.

Positive feedbacks provide sources of growth, explosion, erosion and collapse[8], and consequently will continue to challenge how we approach conservation and restoration. Yelenik and D'Antonio's study highlights the importance of understanding how mechanisms of feedback shift, and how these shifts affect species persistence. Although we will never eliminate surprises, this new perspective will inform where and when we might best intervene in systems to capitalize on their changing dynamics. ∎

Katharine N. Suding *is in the Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, California 94720, USA.*
e-mail: suding@berkeley.edu

1. Scheffer, M. *Critical Transitions in Nature and Society* (Princeton Univ. Press, 2009).
2. Yelenik, S. G. & D'Antonio, C. M. *Nature* **503,** 517–520 (2013).
3. Ehrenfeld, J. G. *Annu. Rev. Ecol. Evol. Syst.* **41,** 59–80 (2010).
4. D'Antonio, C. M. & Vitousek, P. M. *Annu. Rev. Ecol. Syst.* **23,** 63–87 (1992).
5. Mack, M. C. & D'Antonio, C. M. *Ecol. Appl.* **13,** 154–166 (2003).
6. D'Antonio, C. M., Hughes, R. F. & Tunison, J. T. *Ecol. Appl.* **21,** 1617–1628 (2011).
7. Suding, K. N. & Hobbs, R. J. *Trends Ecol. Evol.* **24,** 271–279 (2009).
8. Meadows, D. H. *Thinking in Systems: A Primer* (Chelsea Green, 2008).

PLANETARY SCIENCE

A chunk of ancient Mars

Analysis of a meteorite found in northwest Africa, prosaically named NWA 7533, indicates that it is the first sample of the regolith, or 'soil', of Mars, and is derived from the earliest Martian igneous crust yet identified. SEE LETTER P.513

HARRY Y. MCSWEEN

NASA's decadal survey for planetary science[1] concludes that returning samples of the ancient crust of Mars to Earth ranks among its highest priorities for exploring the Solar System. In this issue, Humayun *et al.*[2] (page 513) describe a Martian meteorite sample already on Earth, albeit without the geological context that samples collected on Mars would have. Nonetheless, it is a revealing discovery.

The meteorite, which is called NWA 7533 (Fig. 1; NWA is an acronym for northwest Africa, where it was found), was part of a celestial rock that broke up during its passage through the atmosphere, producing at least five recovered stones. Another member of this group of stones, NWA 7034, was described previously[3] as a volcanic breccia, which means that it is composed of fragmentary material produced from basaltic lava. Humayun *et al.* have interpreted NWA 7533 — and, by extension, NWA 7034 — as being a regolith breccia. Regolith is the planetary surface layer that is pulverized by meteor impacts (planetary scientists often use the terms 'regolith' and 'soil' interchangeably, which

drives soil scientists mad). Regolith breccias are soils compacted and cemented into rocks by impact-derived melts. Many lunar samples returned by the astronauts of the Apollo missions are regolith breccias.

NWA 7533 contains clasts (fragments) that texturally resemble impact-derived melts in lunar regolith breccias, but with chemical compositions unique to Mars. The compositions of the clasts are nearly identical to those of basaltic rocks and soils analysed by the Spirit rover during its trek through the Gusev Crater on Mars. The high abundance of normally rare elements in the clasts, such as nickel, osmium and iridium, supports the idea that NWA 7533 is a

regolith breccia. Lunar soils are also rich in these elements, because they have been bombarded by chondritic meteors and, over time, become contaminated with their debris. The composition of chondritic meteors is thought to reflect the primordial composition of the terrestrial (rocky) planets before these elements were sequestered into the planets' cores. Contamination by chondritic material also accounts for the high levels of iridium found in strata on Earth from the Cretaceous–Tertiary geological boundary, famously cited as evidence that a meteor impact was responsible for the extinction of the dinosaurs.

The real surprise is the ancient age reported for NWA 7533: 4.4 billion years, demonstrating that this breccia is a sample of the earliest Martian crust. The age was determined by analysing the radioactive-decay products of uranium in zircon crystals, which concentrate this element. Zircon crystals typically form during magma crystallization, and these impervious crystals probably survived pulverization and melting of their host rocks by impacts. The age differs from that previously reported[3] for NWA 7034 (2.1 billion years), an age that was obtained using the decay of radioactive rubidium. The younger age determination, based on analysis of the bulk rock, may represent a mixture of the ages of formation of different components that make up the breccia, or may record some isotopic disturbance that occurred long after the igneous crystallization of the original basaltic rocks. The new, older age implies that a thick Martian crust formed within the first 100 million years or so of the planet's history, coeval with the formation of the Moon's crust.

These new Martian meteorite breccias are fiendishly complex rocks, and forthcoming

Figure 1 | A piece of the NWA 7533 meteorite found in northwest Africa. The width of the stone is 40 mm.

investigations will surely reveal more surprises and conundrums. Detailed studies of the various types of breccia clast, including age dating and analysis of their geochemistry and petrology, could help to unravel the geological record of early Mars.

It has become apparent that Martian meteorites have different chemical compositions from rocks analysed on the planet's surface[4]. Various explanations have been proffered to explain this difference[5,6]. But with the discovery of these latest meteorite breccias, we have a handful of paired meteorites that have the composition of Mars surface rocks, as well as one rock from the Martian surface with a composition like that of the meteorites[7].

Increasingly, the world's meteorite collections are being augmented by finds in hot (northwest Africa) and cold (Antarctica) deserts. Both sources have revealed previously unknown meteorite types, but it is unfortunate that these unique Martian meteorites fell in Morocco rather than on Antarctic ice. The acquisition of meteorites from hot desert countries for research typically depends on the ability to buy them, as opposed to the case with Antarctic meteorites, which are collected, curated and subsampled under nearly pristine conditions and allocated widely and free of charge on the basis of the scientific quality of proposals to study them. But we will gladly accept more samples of Mars from wherever we can get them. ■

Harry Y. McSween *is in the Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, Tennessee 37996–1410, USA.*
e-mail: mcsween@utk.edu

1. National Research Council. *Visions and Voyages for Planetary Science in the Decade 2013–2022* (National Academies Press, 2011).
2. Humayun, M. *et al. Nature* **503,** 513–516 (2013).
3. Agee, C. B. *et al. Science* **339,** 780–785 (2013).
4. McSween, H. Y. Jr, Taylor, G. J. & Wyatt, M. B. *Science* **324,** 736–739 (2009).
5. Tuff, J., Wade, J. & Wood, B. J. *Nature* **498,** 342–345 (2013).
6. Balta, J. B. & McSween, H. Y. Jr *Geology* **41,** 1115–1118 (2013).
7. Zipfel, J. *et al. Meteorit. Planet. Sci.* **46,** 1–20 (2011).

This article was published online on 20 November 2013.

DRUG DISCOVERY

Pocket of opportunity

After three decades of unsuccessful efforts to develop small molecules that neutralize the cancer–causing Ras proteins, an approach has been found that opens up fresh avenues for anticancer research. SEE LETTER P.548

GIDEON BOLLAG & CHAO ZHANG

One-third of all tumours harbour mutations in *RAS* genes[1], but the Ras proteins encoded by these mutant genes have steadfastly eluded targeting by therapeutic agents. On page 548 of this issue, Ostrem *et al.*[2] present perhaps the most promising strategy ever pursued towards developing an anticancer drug that targets mutant Ras proteins. The authors' clever approach was to make compounds that affect a subset of Ras mutations in which a particular amino acid — glycine-12 — in the protein is replaced by another amino acid, cysteine. This kind of mutation, dubbed G12C, is found in a substantial proportion of lung cancers. Because the G12C mutation exists only in tumour cells, drugs that target it could be exquisitely selective, and therefore potentially much less toxic than many current anticancer drugs.

Normal cellular Ras is a small protein that serves as a switch for cell signalling[1]. It binds the nucleotide GTP, hydrolysing it to form another nucleotide, GDP, and so cycles between GTP-bound 'on' and GDP-bound 'off' states. Mutations such as G12C impair GTP hydrolysis and trap Ras in the GTP-bound 'on' state, causing unregulated signalling that can lead to cancer. In the human body, Ras is therefore both friend and foe: the non-mutated protein is the beating heart of cell signalling, but mutated versions are
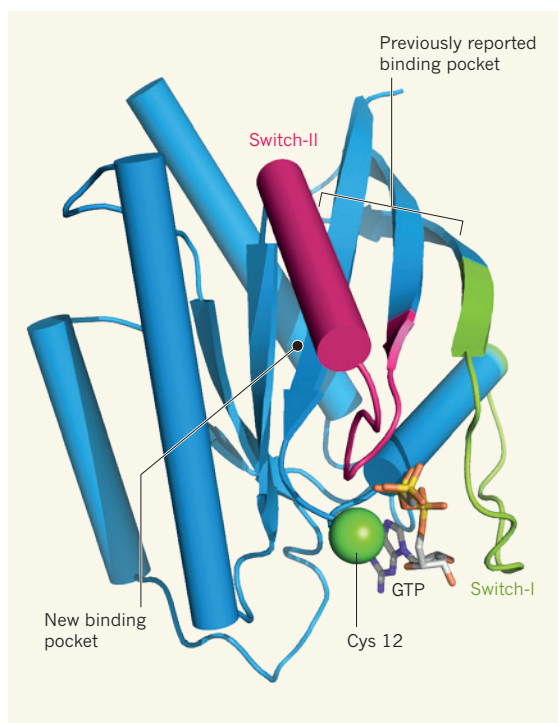

Figure 1 | Binding pockets for Ras inhibitors. In the G12C cancer-causing mutant of the protein Ras, depicted here as a ribbon diagram, a cysteine amino acid (Cys 12) replaces a glycine. Cylindrical sections indicate α-helices; ribbons indicate β-sheets. The protein's substrate, the nucleotide GTP, is bound at the bottom right. Previously discovered Ras inhibitors[5–7] bind to a region between switch-I and switch-II, which are the main regions of Ras that interact with regulators and effector molecules. Ostrem *et al.*[2] have discovered inhibitors that bind irreversibly to Cys 12. This led to the identification of a new binding pocket, which in turn allowed the authors to prepare Ras inhibitors of increased potency.

the villainous masterminds of malignancies.

No drugs that combat Ras-driven human cancers have so far been developed, and the stakes for doing so are high: we cannot win the war on cancer without taming Ras. Accordingly, the US National Cancer Institute this year allocated US$10 million specifically to develop such drugs, and key researchers in the field have committed to making this effort a reality[3].

In fact, medicinal chemists have long attempted to halt unregulated Ras, but were unable to identify small molecules that could access the nucleotide-binding pocket in Ras in order to do so. Drug development at large pharmaceutical companies had therefore been focused on inactivating Ras indirectly by cropping its lipid tail — a feature that it uses to attach itself to cell membranes. This led to the discovery of compounds known as farnesyltransferase inhibitors, which stall an enzyme that is involved in attaching the lipid tail to Ras. But although these compounds were active in animal models, they were ineffective in human patients with cancer, because cancer cells replaced the cropped tail with an alternative one[4]. Earlier this year, biochemists discovered another target[5] to prevent the membrane localization of Ras: a protein called PDEδ. The first PDEδ blockers to be developed inhibit the cancer-causing activity of mutant Ras, but an anxious wait is in store before we know whether this approach will yield effective anticancer drugs.

With recent improvements in drug design guided by protein structures, there is renewed interest in targeting Ras directly. The latest generation of drug developers has thus replaced the previous sledgehammer approach with one that has scalpel-like precision. Within a year, three groups have reported small molecules that directly modulate Ras activity[6–8]. However, the compounds bind weakly to the protein,

because the targeted binding pockets are shallow, and it is unclear whether binding affinity can be improved sufficiently for anticancer applications.

Ostrem *et al.* now report a first volley of Ras inhibitors that work by attacking the cysteine residue of the Ras G12C mutation by means of a thiol (SH) group, forming a disulphide (S–S) bond to the residue's side chain and so providing a foothold on Ras. When the authors obtained an X-ray crystal structure of an inhibitor-tethered protein, they observed a newly exposed pocket that presents new opportunities for drug discovery (Fig. 1). Unfortunately, thiol-based inhibitors are not very 'drug-like' because they are rapidly degraded in cells. However, the researchers have begun to optimize their prototype compounds and have made improved versions carrying 'warheads' that bind irreversibly to cysteine. These compounds are appropriate for biochemical and cellular studies.

Interestingly, Ostrem and colleagues' inhibitors have a preference for the GDP-bound form of Ras, and the authors' biochemical assays show that the compounds prevent the mutant protein from binding GTP. This is desirable behaviour for an anticancer drug because it traps inactive Ras, interrupting signalling through the downstream effectors that cause cancer. Indeed, the researchers find that their partly optimized compounds partially block Ras signalling in cells and exhibit some selectivity for G12C-expressing cells, blocking their proliferation in preference to that of cells that lack a G12C mutation. This selective antiproliferative activity validates the authors' approach and suggests that a truly effective drug might be possible if the compounds can be improved further.

It should be emphasized that even the best of the reported compounds are not suitable for use as drugs. The compounds do not completely block Ras signalling in cells, and it is unclear whether this is a limitation of the overall approach or of the incompletely optimized compounds. Although there are grounds to hope that drugs could be developed that have a substantial therapeutic index (that is, being highly effective with minimal toxicity), there may be a limit to the balance that can be achieved between efficacy and side effects. Moreover, bringing a successful drug to the market will require substantial further investment.

Another caveat concerning Ostrem and colleagues' strategy is that compounds that form irreversible bonds with cysteine are inherently reactive. This might be a problem, because reactive compounds tend to be toxic. However, the idea of anticancer agents that work by irreversibly binding cysteine has gained considerable traction with the recent successes of the drugs afatinib[9] and ibrutinib[10], which act in this way. Nevertheless, Ostrem *et al.* are attempting to further exploit the G12C foothold by designing compounds that no longer require irreversible binding to cysteine, which might lead to compounds that are active against other Ras mutations. This would broaden the population of patients who could benefit from such compounds, but may come at the expense of the drugs' therapeutic index. ∎

Gideon Bollag *and* **Chao Zhang** *are at Plexxikon Inc., Berkeley, California 94710, USA.*
e-mails: gbollag@plexxikon.com;
czhang@plexxikon.com

1. Baines, A. T., Xu, D. & Der, C. J. *Future Med. Chem.* **3**, 1787–1808 (2011).
2. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. *Nature* **503**, 548–551 (2013).
3. Thompson, H. *Nature Med.* **19**, 949–950 (2013).
4. Rowinsky, E. K. *J. Clin. Oncol.* **24**, 2981–2984 (2006).
5. Zimmermann, G. *et al. Nature* **497**, 638–642 (2013).
6. Maurer, T. *et al. Proc. Natl Acad. Sci. USA* **109**, 5299–5304 (2012).
7. Sun, Q. *et al. Angew. Chem. Int. Edn* **51**, 6140–6143 (2012).
8. Shima, F. *et al. Proc. Natl Acad. Sci. USA* **110**, 8182–8187 (2013).
9. Sequist, L. V. *et al. J. Clin. Oncol.* **31**, 3327–3334 (2013).
10. Byrd, J. C. *et al. N. Engl. J. Med.* **369**, 32–42 (2013).

This article was published online on 20 November 2013.

How bacteria choose a lifestyle

In a bacterial population, some cells stay single and motile, whereas others settle down and form chains. A study now investigates the mechanisms that determine these outcomes. SEE ARTICLE P.481

JAMES C. W. LOCKE

Cells can switch identity several times during development. How do they decide to switch? There is much debate about the extent to which identity switching is a cell-autonomous decision as opposed to being driven by environmental signals. In this issue, Norman *et al.*[1] (page 481) take an unusual approach to address this issue. They watch the soil bacterium *Bacillus subtilis* growing in an unchanging environment in which switching cannot be driven by extracellular signals. They focus on a simple switch — the transition from a single-cell swimming (motile) state to a sessile state that allows the bacteria to form a chain. Their findings provide invaluable insight into how an individual cell makes up its mind.

The authors grew *B. subtilis* in a microfluidic device consisting of several channels, each designed to support bacterial growth for days in a constantly replenishing medium that washes away any extracellular signals[2]. The bacterial strains studied express fluorescent 'reporter' proteins for both motile and sessile states, enabling the researchers to quantify the frequency and duration of cell-fate switching events under constant environmental conditions.

Norman and colleagues' detailed and precise characterization of hundreds of switching events reveals a critical difference in the
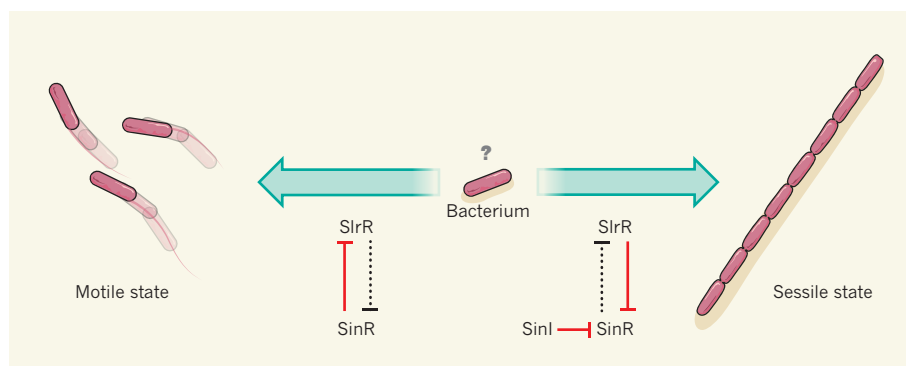
Figure 1 | To be sessile or to swim. Norman *et al.*[1] examine how single cells of the bacterium *Bacillus subtilis* choose between a motile (swimming) state and a sessile state. This choice is controlled by a simple double-negative feedback circuit involving three proteins: SinR represses SlrR, which represses SinR in turn. When SinR dominates, transition to the motile state occurs, whereas when SlrR dominates, the cells become sessile and form chains. A third protein, SinI, can initiate the switch to the sessile state by binding to and inactivating SinR.

transition from the motile to the sessile state and the switch in the other direction. The shift from the motile to the sessile state seems to be completely random and independent of how long the bacterium has been in the motile state. This motile state, therefore, is 'memoryless'.

The switch from the sessile to the motile state, however, is not random and is tightly timed: cells remain in the sessile state for roughly eight generations. The authors suggest that this memory serves a cellular function, ensuring that switching to the motile state, which breaks the chain almost immediately, does not occur too soon or with too much delay, which could result in some chains overflowing with millions of cells. The transition to the sessile state probably represents a trial period of multicellularity, which could be reinforced by environmental signals to commit the cells to forming a biofilm.

Norman and co-workers also explore the molecular mechanism that controls the cell-fate switch. It seems to involve a simple circuit consisting of only three proteins[3] (Fig. 1). Specifically, the protein SinR represses the gene encoding another protein, SlrR; in turn, SlrR binds to and titrates SinR. Thus, these two proteins form a double-negative feedback switch. When SinR wins, the cell enters a motile state; when SinR loses, the cell becomes sessile. The third protein, SinI, affects which outcome wins by binding to, and inactivating, SinR.

The circuit seems to be modular, as the authors find that SinI is responsible for the memoryless entry into the sessile state. Once the bacteria are in the sessile state, however, SinI is no longer relevant, and the memory is set by the SlrR–SinR feedback loop. Such modularity has also been observed in another *B. subtilis* circuit that controls a developmental switch. Under stress conditions, *B. subtilis* can transiently enter a competent state, allowing it to take up external DNA[4]. The core circuitry that controls entry into the competent state has only a few components, similar to the SlrR–SinR–SinI network. The competence circuit is modular because one component regulates the frequency of transitions into the competent state, whereas another component determines how long a cell remains in this state[4].

It is unclear what advantage, if any, such modularity has for the cell. Can having independent control of the initiation and duration of differentiation events enable the cell to adapt to independent selective pressures during evolution? And it remains to be seen whether such modularity is a general feature of circuits that control cell-identity switching.

The authors also raise questions about how the SlrR–SinR–SinI circuit controls cell-fate switching in *B. subtilis*. How noise, or variability, in one of the circuit components drives initiation into a sessile state remains unclear. Although initiation requires SinI, it is not known which circuit component exhibits random fluctuations to drive the random switch

into the sessile state, or how these fluctuations are generated. It would be interesting to test the hypothesis that memory in state switching allows a trial window of multicellularity that is reinforced by environmental signals. One approach could be to examine what effect extending or reducing the memory of the sessile state has on biofilm formation.

Norman and colleagues' SlrR–SinR–SinI circuit joins a growing list of bacterial simple genetic circuits that have been shown to control surprisingly complex cellular dynamics. Such circuits often consist of only three or four proteins but can generate pulses[5], excitable dynamics[4] and robust oscillations[6]. Might simple genetic circuits generate a similar wealth of regulatory dynamics in plants and animals? Results in cows suggest[7] that the concept of memory in state switching could be quite general. Research honoured with the 2013 Ig Nobel prize in probability showed that,

in cows, the standing (motile) state is memoryless, whereas the lying down (sessile) state is timed, just as in *B. subtilis*. ■

James C. W. Locke *is in the Sainsbury Laboratory, University of Cambridge, Cambridge CB2 1LR, UK.*
e-mail: james.locke@slcu.cam.ac.uk

1. Norman, T. M., Lord, N. D., Paulsson, J. & Losick, R. *Nature* **503,** 481–486 (2013).
2. Wang, P. *et al. Curr. Biol.* **20,** 1099–1103 (2010).
3. Chai, Y., Norman, T., Kolter, R. & Losick, R. *Genes Dev.* **24,** 754–765 (2010).
4. Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. *Science* **315,** 1716–1719 (2007).
5. Locke, J. C. W., Young, J. W., Fontes, M., Hernández Jiménez, M. J. & Elowitz, M. B. *Science* **334,** 366–369 (2011).
6. Mackey, S. R., Golden, S. S. & Ditty, J. L. *Adv. Genet.* **74,** 13–53 (2011).
7. Tolkamp, B. J., Haskell, M. J., Langford, F. M., Roberts, D. J. & Morgan, C. A. *Appl. Anim. Behav. Sci.* **124,** 1–10 (2010).

ASTROPHYSICS

Exception tests the rules

Detailed observations of an intermittent ultraluminous X-ray source indicate that its emission is unlikely to be powered by mass accretion onto an intermediate-mass black hole as previously thought. See Letter p.500

K. D. KUNTZ

Ultraluminous X-ray sources (ULXs) are extragalactic sources of X-rays, powered by black holes, that are not coincident with galactic nuclei and have luminosities greater than 10^{39} erg per second — roughly the highest luminosity that stellar-mass black holes, which weigh less than about 30 solar masses, should be able to achieve. Many types of X-ray source could fit this definition, and there has been a recent multiplication of ULX types. On page 500 of this issue, Liu *et al.*[1] use an intermittent ULX in the spiral galaxy M 101 — an object that purists might argue is not a 'real' ULX — to show that several aspects of our understanding of ULXs and, indeed, of black-hole formation, may need to be revised.

The luminosity and spectrum of this object, known as M 101 ULX-1 (Fig. 1), had suggested that it is an intermediate-mass black hole[2]. These black holes have masses in the range of 100 to 1,000 solar masses, and so are larger than black holes formed by the collapse of single massive stars, but smaller than the supermassive black holes that lurk in galactic nuclei. However, Liu and colleagues' radial-velocity measurements of M 101 ULX-1 show that it

is likely to be a black hole with a mass of only 20–30 solar masses. It is not an intermediate-mass black hole, even though commonly used relations between mass, luminosity and temperature imply that it should be.

The Eddington luminosity of an accreting black hole occurs when the pressure of the infalling material that will produce radiation is balanced by the outward radiation pressure. This simple physical argument sets the maximum luminosity for a given black-hole mass, or the minimum mass an accretor must have to produce a given luminosity. High luminosities require high fuelling (accretion) rates, which require the black hole to have a stellar companion to provide the fuel. This, in turn, requires the black hole to have formed from the collapse of a massive star, and to have a mass less than about 30 solar masses.

A typical ULX, at the Eddington luminosity, must have a minimum mass of 100 to 1,000 solar masses, much larger than predicted by our understanding of how stars become black holes. Thus, either our knowledge of the stellar evolution leading to black holes is wrong, or our simple picture of accretion is wrong. Indeed, mechanisms for accretion beyond the Eddington limit have been proposed[3], although their likelihood remains unclear.

Most of the time, M 101 ULX-1 has a luminosity 100-fold less than that of the ULX definition, but it occasionally flares into the ULX regime. Liu *et al.* show that the mass of M 101 ULX-1 is probably 20–30 solar masses, so its luminosity in outburst greatly exceeds the Eddington limit. But because the outbursts are relatively short (less than a week), this super-Eddington luminosity may be understandable. Thus, one might not think this source interesting. However, M 101 ULX-1 contradicts another common understanding about ULXs.

Under some highly simplifying assumptions, standard-accretion models predict the mass of a black hole to be proportional to T^{-4}, where T, which is determined through spectral fitting[4], is the characteristic temperature of a disk formed from the infalling material. More-massive accretors have lower disk temperatures. The disk temperatures of ULXs are lower than those of Galactic black-hole binary systems with measured masses, but at a given disk temperature, ULXs have luminosities that are 100-fold higher[5]. This difference has been taken as tentative evidence that ULXs are indeed much more massive than stellar-mass black holes. However, Liu *et al.* use archival X-ray data to show that M 101 ULX-1, in outburst, has the low temperature expected from a ULX, despite being a stellar-mass black hole. Thus, other ULXs previously thought to be intermediate-mass black holes on the basis of their luminosity and temperature

Figure 1 | The host galaxy of ultraluminous X-ray source M 101 ULX-1. The source is very faint in this image.

may, in fact, be stellar-mass systems.

Liu and colleagues remind us of the need for multi-wavelength data to understand these objects; they used data from a deep year-long monitoring campaign with the Chandra X-ray Observatory, extensive optical imaging by the Hubble Space Telescope, and a major spectroscopic programme with the 8.1-metre Gemini telescope, although some of these data were obtained for other reasons. They have managed to resolve the controversial issue of the nature of the optical counterpart to M 101 ULX-1. This was once thought to be a massive 'B supergiant star', and then a lower-mass 'F star' whose emission was drowned in the optical light from

the accretion disk. But the authors now confirm a previous suggestion[6] that it is a Wolf–Rayet star, an evolved massive star undergoing strong mass loss. The spectroscopic observations also suggest that the system does not have the low metallicity (abundance of elements other than hydrogen and helium) that has come to be associated with ULXs[7].

Thus, although these properties mean that M 101 ULX-1 is not a 'classic' ULX, it was formerly a particularly good intermediate-mass black-hole candidate. It is systems such as these that, when studied in this detail, will allow us to determine the conditions under which super-Eddington accretion occurs, whether ULX luminosities are super-Eddington, and whether intermediate-mass black holes are really needed to explain ULXs. ∎

K. D. Kuntz *is in the Henry A. Rowland Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA.*
e-mail: kuntz@pha.jhu.edu

1. Liu, J.-F., Bregman, J. N., Bai, Y., Justham, S. & Crowther, P. *Nature* **503**, 500–503 (2013).
2. Kong, A., Di Stefano, R. & Yuan, F. *Astrophys. J.* **617**, L49–L52 (2004).
3. Begelman, M. *Astrophys. J.* **551**, 897–906 (2001).
4. Makishima, K. *et al. Astrophys. J.* **535**, 632–643 (2000).
5. Miller, J. M., Fabian, A. C. & Miller, M. C. *Astrophys. J.* **614**, L117–L120 (2004).
6. Liu, J. F. *Astrophys. J.* **704**, 1628–1639 (2009).
7. Prestwich, A. *et al. Astrophys. J.* **769**, 92 (2013).

MOLECULAR BIOLOGY

Antibiotic re–frames decoding

Ketolide antibiotics have been found to induce a ribosomal frameshift — a change in the way that RNA is translated — in bacteria. This promotes the expression of a gene for antibiotic resistance, and may have broader implications.

JOHN F. ATKINS & PAVEL V. BARANOV

Messenger RNAs encode proteins using a sequence of codons: triplets of nucleotides that specify different amino acids. But for a codon sequence to be interpreted, the right reading frame must be set at the start of decoding — translation must begin at the first nucleotide of a codon, rather

than at the second or third. Switching from the correct frame to one of the two alternatives occurs rarely, except at specific places within coding sequences where such frameshifting is programmed. Writing in *Molecular Cell*, Gupta *et al.*[1] report the exciting finding that antibiotics known as ketolides induce frameshifting. This happens during the expression of one of the short coding sequences that often precede

and regulate the expression of genes that encode functional proteins, and is a previously unknown means for such regulation.

Many antibiotics target the ribosome, the nanometre-scale machine in which genetic information is decoded from RNA and translated into proteins in all known living organisms. Certain genes provide bacteria with resistance mechanisms against these antibiotics, such as by encoding proteins that modify ribosomes to prevent antibiotics from binding. But how do these genes sense antibiotics and trigger the synthesis of such defence proteins?

Macrolide antibiotics are sensed in bacteria by their effect on ribosomes that are synthesizing the protein product of a 'leader' sequence. From the perspective of the direction of ribosome movement, the leader RNA lies upstream of the sequence encoding the defence protein. The defence protein is not synthesized in the absence of antibiotics, because its translational start site is hidden in a region of the mRNA that folds in on itself to form a hairpin-like

structure. But when a macrolide antibiotic interacts with a leader-translating ribosome, the ribosome stalls at a particular place on the RNA. This stalling allows the RNA to refold in a different way and exposes the start of the defence sequence; ribosomes that have not yet come into contact with the antibiotic then start to synthesize the defence protein, ultimately enabling the bacterium to resist the antibiotic. This mechanism was worked out during the 1980s[2,3] and was subsequently refined[4], and has long been a classic example of regulation at the mRNA-decoding level.

Gupta and colleagues now report that the sensing of ketolides — which are structurally related to classical macrolide antibiotics — does not occur by the stalling mechanism. Instead, ketolides cause ribosomes to shift reading frame in the same leader sequence (Fig. 1), at a different position from that at which classical macrolides induce stalling. This happens because of a specific perturbation to the standard decoding of codons. The ribosome therefore does not sense its usual 'stop' codon in the original reading frame. Instead, it progresses into a normally non-coding region, preventing the formation of the hairpin structure that usually hinders the translation start site of the defence gene, thus switching on resistance.

In previously known cases in which specific ribosomal frameshifting is used — for instance, to sense levels of an intracellular protein[5] or of small molecules called polyamines[6] and cause a regulatory response

— the synthesis of a frameshift-derived product was the regulated feature. However, Gupta and co-workers' finding is the first reported case in which the functional consequences of ribosomes progressing into 'new territory' are independent of the protein product of that territory. The authors note that the leader sequence for a gene involved in sugar metabolism is similarly organized to the one in their study, and suggest that frameshifting might also be used in expression of that gene.

The translation of short, regulatory leader sequences was first discovered in bacteria, but also occurs in other organisms, including humans. In non-bacterial organisms, such sequences are commonly called upstream open reading frames (uORFs). Unlike in bacteria, and despite their crucial role, uORFs are not given gene symbols and no standard nomenclature exists for their designation. Their importance frequently derives from the motion of ribosomes as they translate uORF sequences, rather than from the synthesis of the uORF-encoded product.

The purpose of uORF translation is often to sense the physiological state of a cell and to modulate expression of the downstream main coding sequence. Molecular biologists' understanding of the ways in which uORF translation modulates expression of the main coding sequence is increasing and, by adding to this knowledge, Gupta and colleagues' discovery will prompt the search for other conceptually similar occurrences. It is becoming apparent that translation of mRNA leader sequences in

mammals and other organisms is extensive. For example, tell-tale 'footprints' of ribosome translation in these regions have been found in abundance[7], and cross-species comparison of sequences has identified highly evolutionarily conserved leader elements[8], suggesting that their functions are vital.

Several genetic disorders in humans are caused by nucleotide substitutions that generate premature stop codons, which terminate protein synthesis before a functional product is made. Therapies in which antibiotic-derived compounds act on human ribosomes so that some of them continue synthesis through a premature stop codon are being clinically evaluated.

Other harmful mutations are insertions or deletions of single nucleotides (or non-multiples of three) that disrupt the reading frame. The effect of these could be relieved by compensatory ribosomal frameshifting. Developing drugs that cause such frameshifting in humans might not be difficult, but their usefulness would be a trade-off between the cell's great capacity to degrade many aberrant products, the amount of almost-normal product generated and how much would be needed for disease amelioration. The success of this approach will depend on several factors, including the proximity of shift-prone sequences to the site of mutational insertion or deletion.

Compounds that induce ribosomal frameshifting could also have antiviral properties, because many viruses use frameshifting for their expression. With retroviruses, for instance, the ratio of the concentration of the frameshift-derived product to that of its counterpart derived from standard decoding is key to viral propagation. Irrespective of the possible medical applications, or of the potential use for controllable switching of protein synthesis in synthetic biology, Gupta and co-workers' discovery raises intriguing questions about translation versatility and the functional diversity of mRNA leader sequences. ∎

John F. Atkins *and* Pavel V. Baranov *are in the School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland.*
J.F.A. *is also in the Department of Human Genetics, University of Utah, Salt Lake City, USA.*
e-mails: j.atkins@ucc.ie; p.baranov@ucc.ie

Figure 1 | Antibiotic-induced frameshifting. Gupta et al.[1] report that ketolide antibiotics trigger ribosomal frameshifting that regulates the expression of a bacterial defence gene. **a**, In the absence of ketolides, a ribosome progresses along a leader region of RNA (grey arrow indicates direction of movement) translating codons until it reaches a 'stop' sequence. The reading frame used by the ribosome is indicated by brackets; red dots represent nucleotides. A downstream 'start' site that triggers the expression of a defence gene is hidden within a hairpin structure, and is therefore inaccessible. **b**, Ketolides cause the ribosome to adopt a new reading frame. It therefore does not recognize the stop codon, and progresses into a normally non-coding region of the RNA, releasing the start site from its hairpin and enabling expression of the defence gene.

1. Gupta, P., Kannan, K., Mankin, A. S. & Vázquez-Laslop, N. *Mol. Cell* http://dx.doi.org/10.1016/j.molcel.2013.10.013 (2013).
2. Horinouchi, S. & Weisblum, B. *Proc. Natl Acad. Sci. USA* **77**, 7079–7083 (1980).
3. Shivakumar, A. G. *et al. Proc. Natl Acad. Sci. USA* **77**, 3903–3907 (1980).
4. Vázquez-Laslop, N. *et al. Mol. Cell* **30**, 190–202 (2008).
5. Craigen, W. J. & Caskey, C. T. *Nature* **322**, 273–275 (1986).
6. Matsufuji, S. *et al. Cell* **80**, 51–60 (1995).
7. Ingolia, N. T. *et al. Science* **324**, 218–223 (2009).
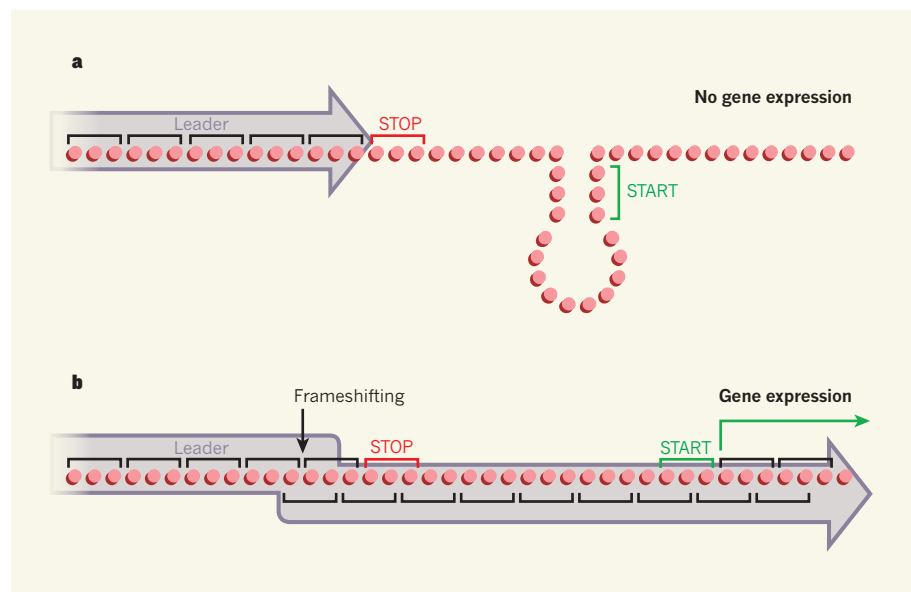8. Ivanov, I. P. *et al. Nucleic Acids Res.* **38**, 353–359 (2010).

ARTICLE

Memory and modularity in cell–fate decision making

Thomas M. Norman[1]*, Nathan D. Lord[1]*, Johan Paulsson[1] & Richard Losick[2]

Genetically identical cells sharing an environment can display markedly different phenotypes. It is often unclear how much of this variation derives from chance, external signals, or attempts by individual cells to exert autonomous phenotypic programs. By observing thousands of cells for hundreds of consecutive generations under constant conditions, we dissect the stochastic decision between a solitary, motile state and a chained, sessile state in *Bacillus subtilis*. We show that the motile state is 'memoryless', exhibiting no autonomous control over the time spent in the state. In contrast, the time spent as connected chains of cells is tightly controlled, enforcing coordination among related cells in the multicellular state. We show that the three–protein regulatory circuit governing the decision is modular, as initiation and maintenance of chaining are genetically separable functions. As stimulation of the same initiating pathway triggers biofilm formation, we argue that autonomous timing allows a trial commitment to multicellularity that external signals could extend.

Cell-fate decisions often result from explicit extracellular triggers[1–3]. It is now appreciated that internal stochastic fluctuations[4–10] can also drive a cell to switch fates even in the apparent absence of external signals[11–17]. Neighbouring cells in the developing gonad of *Caenorhabditis elegans* compete to become ventral uterine or anchor cells[18], and subpopulations of growing *Escherichia coli* cells probabilistically enter a quiescent, antibiotic-resistant state[14,19]. But whether occurring in the body of a nematode or in shaking culture, these decisions take place against a backdrop of environmental change driven by continued growth. With rising interest in quantitative properties of gene networks[20,21], one central question is how much of a cell's behaviour can be attributed to the environment and how much to the internal program, that is, the behaviour the network would implement were the environment fixed.

A prototypical situation arises in the conversion of bacteria from free-living, planktonic cells into sessile, multicellular communities known as biofilms[22,23]. Like many complex fates, biofilm formation is a product not just of a cell's individual behaviour, but also of reinforcement by environmental cues created by nutrient depletion, the production of matrix[24], quorum sensing[25], and hypoxia[26]. Here we use a microfluidic device to investigate the earliest stages of multicellular growth by the soil bacterium *Bacillus subtilis*. Our approach removes confounding environmental influences while allowing for high-throughput quantitative imaging, thereby revealing the cell's internal programs of development.

B. subtilis provides a natural model system for decision making. During the exponential phase of growth, it exists in two states: as individual, motile cells and as long, connected chains of sessile cells[27]. Switching between these states has been thought of as a bet-hedging strategy[28–30], with motile cells acting as foragers and chains representing periodic attempts to settle down and start a colony. At the heart of the decision is a simple three-protein network between SinI, SinR and SlrR (refs 31, 32). Commitment to each state is controlled by a double-negative feedback loop in which SinR represses the *slrR* gene, and SlrR binds to and titrates SinR (Fig. 1a). Motility corresponds to the SlrR^low state in which SinR represses the gene for SlrR and other chaining-associated genes. Chaining occurs during the SlrR^high state in which

SlrR forms a complex with SinR, both titrating its activity against chaining genes and redirecting it to repress motility-associated genes[33]. Although both states are present during exponential growth, the chained state is strongly reinforced during biofilm formation by further antagonism of SinR by SinI, which is produced in response to environmental signals[34,35]. This three-gene network thus supports a two-state process of decision making that can be influenced by environmental signals.

Visualizing fate switching in real time

Microfluidic systems that allow individual cells to be imaged over time as the growth medium is replenished provide an excellent opportunity to examine autonomous developmental programs. Extracellular signalling is removed, and cells cannot accumulate and starve themselves. Building on previous studies[14,36–40], we constructed microfluidic channels from polydimethylsiloxane (PDMS, Fig. 1b) that were sized to accommodate chains of *B. subtilis* (75 μm long and 1.6 μm wide). A unique feature of our design is the shallow side channels that surround the cells, creating a 'bath' of medium that enables efficient feeding over long length scales[41]. The channels are closed on one end, and on the other they empty into a feeding channel that supplies fresh medium (by diffusion) and washes away excess cells as they are pushed out by growth. To prevent cells from swimming out of the channels, the ability of the flagellum to generate force was disrupted through a straight flagellum mutation[42].

Only motile cells expressed the flagellin gene (Supplementary Video 1) as visualized with a P_{hag}-mKate2 reporter (coloured green), and only chains expressed matrix genes as visualized with a P_{tapA}-cfp reporter (coloured red). We therefore used these reporters as proxies for the corresponding phenotypic states. *B. subtilis* interconverted between the motile and chained states while growing in the channels (Fig. 1c and Supplementary Video 2), leading to anticorrelated flagellin and matrix gene expression. In keeping with the premise that the chains had switched to the SlrR^high state, imaging of *slrR* (visualized with a P*slrR-mKate2* reporter, artificially coloured green) and matrix coexpression

[1]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA.
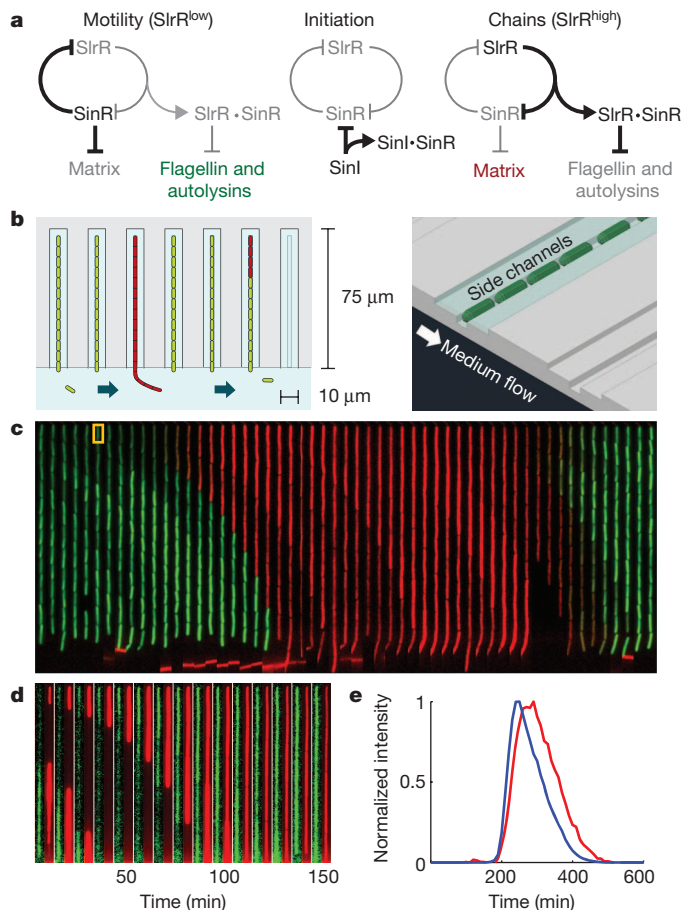*These authors contributed equally to this work.

Figure 1 | Tracking cell-fate switching in *Bacillus subtilis*. a, Genetic logic governing the motile and chained states. **b**, Top and isometric schematics of microfluidic channels in which individual bacteria are held. Channels connect to a larger channel through which medium is continuously replaced and excess cells are washed away. **c**, Kymograph showing a single cell (highlighted in yellow) of strain TMN690 (P_{hag}-*gfp* P_{tapA}-*mKate2 hagA233V*) transitioning from motile growth (marked in green by expression of a P_{hag}-*gfp* reporter for flagellin) to chained growth (marked in red by expression of a P_{tapA}-*mKate2* reporter for matrix expression). Frames are taken 10 min apart. **d**, Kymograph showing co-expression of matrix and *slrR* reporters in TMN1180 cells (P_{tapA}-*cfp* P*slrR*-*mKate2 hagA233V*). **e**, Average co-expression profiles of matrix (blue curve) and *slrR* (red curve) reporter expression in chains (TMN1180, 25 events).

Figure 2 | Dynamics of cell-fate switching. This figure examines chaining in strain TMN1157 (P_{hag}-*mKate2* P_{tapA}-*cfp hagA233V*). **a**, The uppermost cell's fate was tracked in each channel, yielding traces of flagellin (P_{hag}-*mKate2*, green curve) and matrix (P_{tapA}-*cfp*, red curve) reporter expression. Five chaining events are shaded. AU, arbitrary units. **b**, Correlation between subsequent residence times in the motile state. **c**, Schematic of ageing curves. Memoryless switching (blue dashed curve) between states gives rise to horizontal curves, whereas deterministic timers (green dashed curve) create curves descending with slope −1 from the average duration of the state $\langle T \rangle$. Many other mechanisms are bounded by these extremes (Supplementary Information): for example, progression through a series of discrete, exponentially distributed steps yields the grey curve. **d**, Distribution of motility periods (307 events). Red curve shows exponential fit. Inset shows log transformed cumulative distribution function of motility period duration (black curve) and the exponential fit (red curve). **e**, The ageing curve for the motile state (black curve) is compared to the expectation for memoryless switching adjusted for undersampling of long motility periods (blue dashed curve; see Supplementary Information) and that for a timer (green dashed curve). **f**, Distribution of chain durations (440 events). **g**, Ageing curves for chains (blue curve) in cells wild type for *slrR* (TMN1157) and pulses (red curve) in *slrR* mutant cells (TMN1158, which is TMN1157 mutated for *slrR*). All qualitative features of distributions were replicated in at least three separate experiments and quantitative parameters in at least two.

revealed that *slrR* was expressed in chains (Fig. 1d), and that matrix and *slrR* expression were tightly correlated in time (Fig. 1e).

Several million cell divisions were imaged, but we only report data for the fates of the uppermost cell in each channel, as these could be monitored throughout the experiment without being washed away (Fig. 2a and Supplementary Video 2). We thus tracked the histories of thousands of individual bacteria through ~300 generations of growth. To define more precisely the motile and chained states, we found thresholds on the matrix reporter that coincided with onset of matrix expression and the subsequent return of motility, but similar results were obtained for a range of thresholds (Extended Data Fig. 1). All measured properties remained constant in time and across the device: a generation time of ~27 min was sustained for as long as 7 days (Extended Data Fig. 2), chaining occurred at a uniform rate (Extended Data Fig. 3), and within each lineage there was no correlation between the lengths of successive visits to the motile state (Fig. 2b) or the chained state (Extended Data Fig. 4). The switching behaviour was thus homogeneous throughout the device and experiment duration, reflecting a stochastic process at steady state. With the influence of environmental
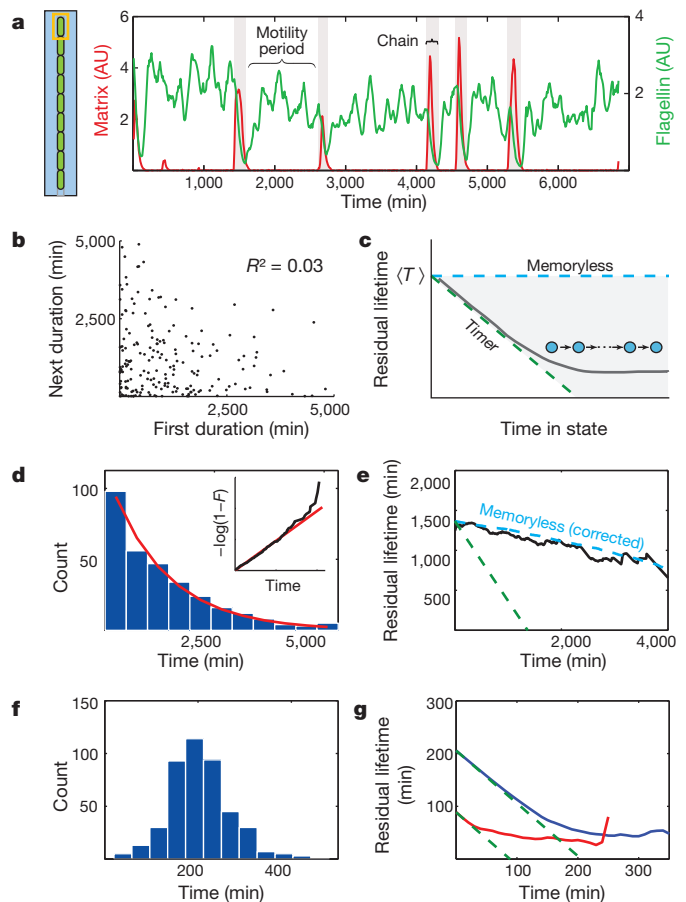
changes removed, we next set out to characterize the autonomous motility and chaining programs.

Memoryless motility and timed chaining

We monitored transitions between motile and chained states to determine whether cells exercise temporal control, or if they exit states independently of their history. The latter memoryless behaviour would imply exponentially distributed residence times between events and thus a coefficient of variation (standard deviation divided by mean) in residence times of $CV = 100\%$, whereas other switching mechanisms could exploit history-dependence to produce narrower distributions. We

further quantified history dependence by asking how each state 'ages', as measured by mean residual lifetime curves, that is, the expected time left in a state given that the system is still there, as a function of time. Memorylessness produces horizontal ageing curves (blue line in Fig. 2c) whereas perfect timing produces linearly decreasing curves with a slope of -1 (green curve in Fig. 2c)[43].

The distribution of residence times in the motile state was almost perfectly exponential with a mean of ~81 generations (~36 h) and CV ≈ 100% (Fig. 2d) after correcting for the length of the experiment (see Supplementary Information). The ageing curve also conformed to the expectation for an exponential random variable (Fig. 2e), and we observed no correlations between the residence times of successive events. Thus, despite the complex underlying circuit, cells decide to chain according to the simplest possible switching scheme: a motile cell does not 'remember' when it last chained, and the probability of chaining is the same whether the cell has been motile for one generation or hundreds of generations.

Chains displayed a radically different behaviour. The residence time distribution was sharply peaked at a mean of 7.6 generations and had a 28% relative standard deviation (Fig. 2f), resembling a gamma distribution with a shape parameter of 13 and with an ageing curve prototypical of tight timing before eventually flattening out (Fig. 2g). Thus, whereas motile cells set long average residence times and allow widely variable commitments, chains instead orchestrate briefer, tightly timed transitions. This difference makes teleological sense given their respective lifestyles. As motile cells grow as individuals, their properties are insensitive to how long they remain motile, leaving no obvious reason to keep track of the residence time. In contrast, any decision that depends on coordination among progeny will require some degree of memory. Chains have strong incentives to preclude both very short and very long commitments. The chained phenotype accumulates over time, where chaining for T generations produces chains of length 2^T. Relatively small differences in T then translate into great differences in chain length. Memoryless exit from the chained state would in fact have extreme consequences, where many chains would break down almost instantaneously whereas others could contain millions of cells. The narrow time distribution guarantees a minimum chain length while preventing a high fraction of cells from effectively entering the chained state irreversibly.

Memoryless initiation from noisy antagonism

Slow and memoryless switching can arise from positive feedback loops, in which rare fluctuations allow the system to break out of the basin of attraction of each stable point[44]. Indeed, one of the key features of the motility-chaining decision network is the SinR–SlrR double negative feedback loop. As expected, mutating *slrR* eliminated chaining: over the course of a 6-day experiment, we saw sustained high expression of flagellin in all cells and observed no morphological evidence of cells growing as connected groups. However, our sensitive time-lapse microscopy allowed us to detect exceedingly rare and weak expression signals, showing that an *slrR* mutant exhibited small and infrequent bursts of matrix expression (Fig. 3a and Supplementary Video 3). We refer to these events as pulses, to distinguish them from chains that pair high matrix expression with repression of flagellin. We note that they also appear in the wild-type data, but fail to trigger expression of *slrR* (Extended Data Fig. 5). Notably, the residence times between subsequent initiation attempts, whether resulting in chains or pulses, followed indistinguishable exponential distributions for wild-type cells and the *slrR* mutant (Fig. 3b). Removal of SlrR thus abolished the chaining phenotype, but left the memoryless process of initiation perfectly intact.

Having determined that initiation arose from a factor upstream of the feedback loop, we examined the SinI protein that antagonizes SinR during biofilm formation. SinI was sufficient to drive chaining, as cells containing an IPTG (isopropyl β-D-1-thiogalactopyranoside)-inducible *sinI* gene rapidly chained upon induction. It was also necessary: cells
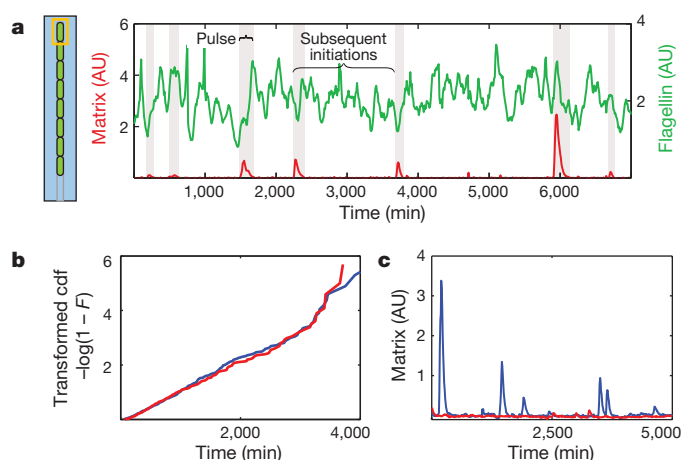

Figure 3 | Memoryless initiation of chaining. a, An example trace of flagellin (P_{hag}-*mKate2*, green curve) and matrix (P_{tapA}-*cfp*, red curve) reporter expression from *slrR* mutant cells (TMN1158). Seven matrix pulses are shaded. AU, arbitrary units. **b**, Log transformed cumulative distribution functions of times between subsequent initiations (of pulses or chains) in cells from wild type (blue curve, TMN1157, 399 events) or mutant for *slrR* (red curve, TMN1158, 296 events) strains. Plotted this way, exponential distributions yield straight lines. This result separately reproduced in a strain with different fluorescent reporter proteins. **c**, Example matrix expression traces in *slrR* mutant cells (blue curve, TMN1158), and in *slrR* mutant cells further deleted for the initiator (*sinI*) (red curve, TMN1198).

mutant for SinI did not chain, and pulses were absent in cells doubly mutant for SinI and SlrR (red curve in Fig. 3c). These results suggest that noisy antagonism of SinR by SinI drives spontaneous chaining in a way that is quantitatively independent of the SlrR feedback control system, as discussed below.

To test how cells control the duration of the chained state, we briefly switched (10 min) on the inducible *sinI* gene to provide a defined initiating signal (Fig. 4a and Supplementary Video 4). Notably, the ageing behaviour of the resulting chains was virtually identical to that of spontaneously occurring chains (see Figs 4b and 2g). Even switching on SinI synthesis a second time in cells that had started to revert from chaining (3 h after first pulse) or using a longer initiating signal led to no increase in the average duration of the resulting chains (Extended Data Fig. 6). The chained state is thus stereotyped: once a signal to chain is registered, the same program is executed in a way that is independent of the nature of the initiating signal or of the history of the cell. This tight timing is an intrinsic property of the SinR–SlrR feedback loop rather than the initiating event, as the spontaneous pulses seen in *slrR* mutant cells showed little evidence of temporal organization (red curve in Fig. 2g). Furthermore, chains lasted longer than pulses under both spontaneous and induced conditions (Figs 4c, d), suggesting that the feedback loop coordinated action after the initiating signal had faded. Indeed, adding an additional copy of *slrR* to strengthen feedback led to longer chaining events (Extended Data Fig. 7). Thus, we again see network modularity[45]: just as the SinR–SlrR feedback loop did not affect the initiation of chaining, the duration of the chained state was independent of the initiation process.

To dissect how cells time their exit from the chained state, we analysed the temporal pattern of gene expression during hundreds of chaining events. Examining the rate of gene expression in these traces (Methods) revealed two distinct phases: a build-up phase of matrix expression was followed by a pure dilution phase when expression was negligible and levels exponentially decreased due to growth (Fig. 4e). Motility then reinitiated once levels fell below a threshold. The two phases were approximately equal in length, with the duration of the dilution phase more narrowly distributed than the build-up phase ($CV_{build-up} = 0.44$, $CV_{dilution} = 0.23$).
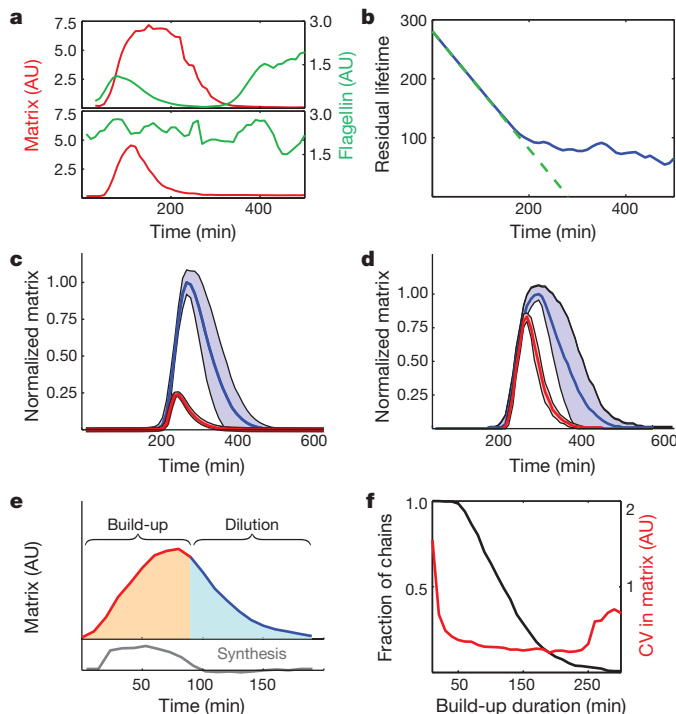
Figure 4 | SlrR executes a stereotyped chaining program. a, Example matrix and flagellin traces from strains where chaining (top panel, TMN1195 = P_{hag}-$mKate2$ P_{tapA}-cfp $hagA233V$ P_{spank}-$sinI$) or pulsing (bottom panel, TMN1196, which is TMN1195 mutated for $slrR$) were inducible by addition of IPTG. AU, arbitrary units. **b,** Ageing curve for induced chains is shown (177 events). Green dashed curve shows expectation for a timer. **c,** Average matrix expression profiles for chains arising spontaneously (blue curve, TMN1157, 198 events) and pulses arising spontaneously in $slrR$ mutant cells (red curve, TMN1158, 278 events). Shaded regions denote ±1 standard deviation. Average profiles are scaled to reflect the average height difference between chains and pulses. **d,** The same analysis for chains (blue curve, TMN1195, 26 events) and pulses (red curve, TMN1196, 42 events) induced by addition of IPTG. **e,** Matrix expression during chaining naturally breaks down into a build-up phase (red curve), where synthesis of new proteins dominates, and a subsequent dilution phase (blue curve). Grey curve shows the calculated synthesis rate (see Supplementary Information) used to call the two phases. **f,** Long build-up phases reduce noise in matrix expression by time averaging. The plot shows the fraction of chains achieving a build-up phase of a given duration (black curve) and the variability in matrix expression of those chains (red curve). Similar results have been obtained in three replicate experiments.

Expression rates in the build-up phase varied substantially between chains at any given time (Extended Data Fig. 8), but also over time in any given cell. By ensuring that each chain committed to an extended build-up phase, SlrR allowed cells to effectively 'time-average'[4] over such noisy expression rates as the total amount of accumulated protein reflected the average of a long history of expression. Because the build-up phase was longer than the correlation time of the random expression process, the variability between chains in matrix gene expression decreased substantially as the build-up phase progressed (Fig. 4f).

Variation in the outcome of the build-up phase meant that cells with higher expression require more time to dilute, but the mechanism of dilution naturally suppresses this heterogeneity. First, because the dilution rate is set by cell growth rather than by a noisy reaction network, dilution can potentially extend the time spent in the state without adding heterogeneity. Indeed, we found that the dilution phase proceeded largely deterministically: the reporter's intensity at the onset of dilution precisely predicted the exit time, and the trajectories were well described by exponential decay (Extended Data Fig. 9). The threshold marking the end of dilution and entrance into the motile state thus

seemed high enough that random segregation of molecules between daughter cells at low numbers[10,46] was made irrelevant. Second, the exponential nature of dilution—reducing levels twofold every generation—further tightened control by making the timing robust to heterogeneity in the initial level of protein. Specifically, the time spent diluting then depends logarithmically rather than linearly on the initial amount. Cells that, by chance, have much more or less protein initially, will then vary marginally in the time spent diluting. Indeed, the 30% deviations in matrix abundance at the onset of dilution was reduced to a 23% deviation in the dilution time, closely following the expectation from a noise-free exponential dilution process (Supplementary Information). Thus, by extending the build-up phase in chains, SlrR is responsible for translating widely variable initiating signals into a precisely timed pattern of gene expression.

Memory enables multicellular cooperation

The choice between motility and multicellularity is central to the lives of many bacteria, as cells must relinquish their autonomy to benefit from living together[22,23]. The chaining program may underlie the earliest steps of multicellularity: by coordinating behaviour across many generations, the tight timing provided by SlrR enforces cooperation among the progeny of a cell that initiates a new sessile community. The long-term commitment to chaining seen during biofilm formation could in turn rely on continued initiation or on feedback mechanisms that lock cells into the multicellular state. Although we saw no evidence that SlrR feedback could provide the requisite commitment, the initiator SinI is indeed strongly expressed both in response to desirable niches (for example, plant polysaccharides)[47] and growth-related stresses (for example, starvation or hypoxia)[24,26]. Our results show that different environmental signals are channelled into the same robust chaining behaviour, and cessation of the stimulus ultimately leads to coordinated exit. Maintenance is thus contingent on continued stimulation, but even small signals will suffice to renew commitment. The role of SlrR feedback may thus be to provide a well-defined 'trial period' of multicellular growth, the continuation of which is periodically re-evaluated.

Regulation of chaining weaves together stochastic gene expression, transcriptional feedback and post-translational regulation. Any quantitative property of the decision could therefore have been a product of several factors acting together. Yet observation of thousands of chaining decisions free from environmental influences revealed a modular network that separates initiation from control of the residence time; eliminating one function leaves the other intact in quantitative detail, allowing the overall behaviour to be explained in terms of these two pieces. This type of excitable dynamics, in which the system is randomly kicked out of a stable state but returns after a well-defined excursion, is often explained in terms of linked feedback loops, and has been implicated in other *B. subtilis* decision networks[16,17]. In this case, however, an exceedingly simple alternative mechanism may explain most of the behaviour. SinR and SinI are known to form an inactive complex with binding constants in the nanomolar range[48]. Because more SinR is produced than SinI, SinR typically titrates out all free SinI molecules, thereby acting as a buffer against small fluctuations. However, a rare persistent accumulation of SinI levels transiently reverses the roles, leading to a buffering pool of free SinI instead. This mechanism can generate long periods of virtually no free SinI (corresponding to the motile state) followed by long stretches of SinI dominance, which induces chaining. The memory in the chained state is in turn largely explained by the production-dilution mechanism above, in which feedback could have a role in narrowing the probability distribution of time spent producing matrix proteins.

Other systems may also display memory and memorylessness for the times spent sessile and motile, respectively[49], but we suspect any broader principles will follow from the sensitivity of a phenotype to the time spent in the state. Decisions that aim only to set the occupancy of a particular state[14,15,19] do not require explicit timing, and

may therefore randomize commitments with memoryless switching. In contrast, when the effectiveness of a cell-fate choice is tied to population size[50], timed decision making could again be used to ensure cooperation among progeny. In metazoans, stochastic cell-fate decisions are often stabilized after the fact by lateral inhibition[18]. Timing the adopted state could provide an initial window of commitment to give extracellular feedback time to take hold. Our approach—observing the cell's intrinsic dynamics while keeping everything else static for extended periods of time—may reveal that many complex developmental choices can be explained by surprisingly simple dynamical principles in individual cells.

METHODS SUMMARY

Strains were grown to high density and loaded into freshly cast and bonded microfluidic chips. A straight flagellum mutation in all strains (hagA233V) prevents the flagellum from generating force so that motile cells cannot swim out of the channels. Fresh LB medium was continuously supplied using syringe pumps, and an automated fluorescence microscope maintained at 37 °C was used to image cells every 10 min. When needed, 10-min pulses of 100 μM IPTG were used to induce chaining. The top cell in each channel was segmented (Extended Data Fig. 10) and its fluorescence was quantified using a Matlab analysis pipeline. Resulting reporter traces were used to produce residence time distributions by finding thresholds on the matrix reporter that identified when the signal was first distinguishable from background, and when motility reporter expression subsequently returned. The time between these two points was defined as the duration of a chain or pulse, and the time between subsequent peaks was defined as the time spent motile. The log transform of a cumulative distribution function $F(t)$ is $-\log[1 - F(t)]$, which for exponential distributions yields a straight line. For a distribution of times T, the ageing curve is $m(t) = E[T - t \mid T > t]$. Average chain and pulse profiles were compiled by normalizing each peak height to 1, registering the leading edges and averaging the aligned peaks. This normalization removes variation due to peak height but leaves variation due to timing behaviour intact. Chain 'build-up' and 'dilution' phases were identified by fitting matrix reporter traces to a kinetic model and extracting expression rates at each point. The build-up phase extends from beginning of the chain to the point where the dilution rate is fivefold larger than the matrix expression rate, and the dilution phase comprises the remainder of the chain.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Ferrell, J. E. Jr & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. Science 280, 895–898 (1998).
2. Wehrli, M. & Tomlinson, A. Epithelial planar polarity in the developing Drosophila eye. Development 121, 2451–2459 (1995).
3. Long, T. et al. Quantifying the integration of quorum-sensing signals with single-cell resolution. PLoS Biol. 7, e68 (2009).
4. Paulsson, J. Summing up the noise in gene networks. Nature 427, 415–418 (2004).
5. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. Nature Genet. 31, 69–73 (2002).
6. Newman, J. R. et al. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441, 840–846 (2006).
7. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. Gene regulation at the single-cell level. Science 307, 1962–1965 (2005).
8. Bar-Even, A. et al. Noise in protein expression scales with natural protein abundance. Nature Genet. 38, 636–643 (2006).
9. Taniguchi, Y. et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538 (2010).
10. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. Science 297, 1183–1186 (2002).
11. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. Nature 467, 167–173 (2010).
12. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135, 216–226 (2008).
13. Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. Nature 435, 228–232 (2005).
14. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. Science 305, 1622–1625 (2004).
15. Huang, G. et al. Bistable expression of WOR1, a master regulator of white-opaque switching in Candida albicans. Proc. Natl Acad. Sci. USA 103, 12813–12818 (2006).
16. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545–550 (2006).
17. Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. Tunability and noise dependence in differentiation dynamics. Science 315, 1716–1719 (2007).
18. Greenwald, I. LIN-12/Notch signaling: lessons from worms and flies. Genes Dev. 12, 1751–1762 (1998).
19. Lewis, K. Persister cells, dormancy and infectious disease. Nature Rev. Microbiol. 5, 48–56 (2007).
20. Pedraza, J. M. & van Oudenaarden, A. Noise propagation in gene networks. Science 307, 1965–1969 (2005).
21. Blake, W. J., KÆrn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. Nature 422, 633–637 (2003).
22. Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. Sticking together: building a biofilm the Bacillus subtilis way. Nature Rev. Microbiol. 11, 157–168 (2013).
23. Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the natural environment to infectious diseases. Nature Rev. Microbiol. 2, 95–108 (2004).
24. Rubinstein, S. M. et al. Osmotic pressure can regulate matrix gene expression in Bacillus subtilis. Mol. Microbiol. 86, 426–436 (2012).
25. Davies, D. G. et al. The involvement of cell-to-cell signals in the development of a bacterial biofilm. Science 280, 295–298 (1998).
26. Kolodkin-Gal, I. et al. Respiration control of multicellularity in Bacillus subtilis by a complex of the cytochrome chain with a membrane-embedded histidine kinase. Genes Dev. 27, 887–899 (2013).
27. Kearns, D. B. & Losick, R. Cell population heterogeneity during growth of Bacillus subtilis. Genes Dev. 19, 3083–3094 (2005).
28. Acar, M., Mettetal, J. T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. Nature Genet. 40, 471–475 (2008).
29. Dubnau, D. & Losick, R. Bistability in bacteria. Mol. Microbiol. 61, 564–572 (2006).
30. Veening, J. W. et al. Bet-hedging and epigenetic inheritance in bacterial cell development. Proc. Natl Acad. Sci. USA 105, 4393–4398 (2008).
31. Chai, Y., Norman, T., Kolter, R. & Losick, R. An epigenetic switch governing daughter cell separation in Bacillus subtilis. Genes Dev. 24, 754–765 (2010).
32. Chai, Y., Kolter, R. & Losick, R. Reversal of an epigenetic switch governing cell chaining in Bacillus subtilis by protein instability. Mol. Microbiol. 78, 218–229 (2010).
33. Chen, R., Guttenplan, S. B., Blair, K. M. & Kearns, D. B. Role of the σD-dependent autolysins in Bacillus subtilis population heterogeneity. J. Bacteriol. 191, 5775–5784 (2009).
34. Kearns, D. B., Chu, F., Branda, S. S., Kolter, R. & Losick, R. A master regulator for biofilm formation by Bacillus subtilis. Mol. Microbiol. 55, 739–749 (2005).
35. Bai, U., Mandic-Mulec, I. & Smith, I. SinI modulates the activity of SinR, a developmental switch protein of Bacillus subtilis, by protein-protein interaction. Genes Dev. 7, 139–148 (1993).
36. Teng, S. W., Mukherji, S., Moffitt, J. R., de Buyl, S. & O'Shea, E. K. Robust circadian oscillations in growing cyanobacteria require transcriptional feedback. Science 340, 737–740 (2013).
37. Wang, P. et al. Robust growth of Escherichia coli. Curr. Biol. 20, 1099–1103 (2010).
38. Moffitt, J. R., Lee, J. B. & Cluzel, P. The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. Lab Chip 12, 1487–1494 (2012).
39. Ullman, G. et al. High-throughput gene expression analysis at the level of single proteins using a microfluidic turbidostat and automated cell tracking. Phil. Trans. R. Soc. Lond. B 368, 20120025 (2012).
40. Robert, L. et al. Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch. Mol. Syst. Biol. 6, 357 (2010).
41. Mather, W., Mondragon-Palomino, O., Danino, T., Hasty, J. & Tsimring, L. S. Streaming instability in growing cell populations. Phys. Rev. Lett. 104, 208101 (2010).
42. DeLange, R. J., Chang, J., Shaper, J. & Glazer, A. Amino acid sequence of flagellin of Bacillus subtilis 168. III. Tryptic peptides, N-bromosuccinimide peptides, and the complete amino acid sequence. J. Biol. Chem. 251, 705–711 (1976).
43. Muth, E. J. Memory as a property of probability distributions. IEEE Trans. Reliab. R-29, 160–165 (1980).
44. Aldous, D. in Probability Approximations via the Poisson Clumping Heuristic (Springer, New York, 1989).
45. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. Nature 402 (suppl.) C47–C52 (1999).
46. Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. Nature Genet. 43, 95–100 (2011).
47. Beauregard, P. B., Chai, Y., Vlamakis, H., Losick, R. & Kolter, R. Bacillus subtilis biofilm induction by plant polysaccharides. Proc. Natl Acad. Sci. USA 110, E1621–E1630 (2013).
48. Newman, J. A., Rodrigues, C. & Lewis, R. J. Molecular basis of the activity of SinR protein, the master regulator of biofilm formation in Bacillus subtilis. J. Biol. Chem. 288, 10766–10778 (2013).
49. Tolkamp, B. J., Haskell, M. J., Langford, F. M., Roberts, D. J. & Morgan, C. A. Are cows more likely to lie down the longer they stand? Appl. Anim. Behav. Sci. 124, 1–10 (2010).
50. Diard, M. et al. Stabilization of cooperative virulence by the expression of an avirulent phenotype. Nature 494, 353–356 (2013).

Author Contributions T.M.N. and N.D.L. designed and fabricated the microfluidic device, cloned strains and collected the data. All authors were involved in conceiving the study, analysing results and writing the paper. J.P. and R.L. are corresponding authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P. (Johan_Paulsson@hms.harvard.edu) or R.L. (losick@mcb.harvard.edu).

METHODS

Strain construction. All strains were derived from *Bacillus subtilis* NCIB3610 using standard molecular biology techniques. Strain genotypes, full construction details and a list of primer sequences are provided in the Supplementary Information. To prevent motile cells from swimming out of the channels, all strains bore a *hagA233V* straight flagellum mutation, which impairs the ability of the flagellum to generate force while leaving its construction intact[42].

Microfluidic device fabrication. The master mould for the device was fabricated in four layers by ultraviolet photolithography using standard methods (for detailed protocol, see Supplementary Information). For each layer, Shipley or SU-8 (Microchem) photoresist was applied to a silicon wafer by spin coating to appropriate thickness (corresponding to the channel height) and patterns were then created by exposing the uncured photoresist to ultraviolet light through custom quartz-chrome photomasks (Toppan Inc.).

Microfluidic devices were fabricated by moulding channel features into a polydimethylsiloxane (PDMS) slab and then bonding that slab to a glass coverslip. To produce the slab, dimethyl siloxane monomer (Sylgard 184) was mixed in a 5:1 ratio with curing agent, poured onto the silicon wafer master, degassed under vacuum, and cured at 65 °C overnight. Holes to connect the feeding channels to the external tubing used for medium perfusion were then introduced using a biopsy punch, and individual chips were cut and bonded onto KOH-cleaned cover slips using oxygen plasma treatment the day of the experiment. Bonded chips were baked at 65 °C for at least an hour before use.

Cell preparation and device loading. Immediately before use, the microfluidic device was passivated with a 10 mg ml^{-1} solution of bovine serum albumin (BSA). *B. subtilis* cells were grown to late stationary phase in LB to decrease their size and thus increase efficiency of loading. They were then passed through a 5 μm filter (Pall Acrodisc) to remove chains, concentrated by centrifugation, and injected into the feeding channel. The chip was mounted on a custom-machined platform that could be inserted into a standard bench-top centrifuge, and cells were forced into the cell channels by centrifugation. Syringes containing LB medium with 0.1 mg ml^{-1} BSA were connected to the device using Tygon tubing (VWR), and were pumped at a flow rate of 3 μl min^{-1} using syringe pumps (New Era Pump Systems). BSA was provided as a lubricant to prevent cells (and chains in particular) from adhering to the surface of the main feeding conduit as they are pushed out of the device.

Microscopy and image acquisition. The microfluidic device was mounted on a fluorescence microscope immediately after loading. We used a Nikon Eclipse Ti inverted microscope equipped with an Orca R2 (Hamamatsu) camera, a ×60 Plan Apo oil objective (NA 1.4, Nikon), an automated stage (Ludl), and a Lumencor SOLA fluorescent illumination system. Image acquisition was performed using Matlab scripts interfacing with μManager[51]. The microscope was encased in a custom-built incubator that maintained it at 37 °C throughout the experiment. The following filter sets were used for acquisition: GFP (Semrock GFP-1828A), mKate2 (Semrock mCherry-B), CFP (Semrock CFP-2432C), YFP (Semrock YFP-2427B). The *slrR/tapA* co-expression experiment was performed on an almost identically configured microscope that instead had a Lumencor SPECTRA fluorescent illumination system. Exposures were done at very low illumination intensities with 2 × 2 binning (CCD chip dimension of 1,344 × 1,024 pixels, pixel size of 6.45 μm × 6.45 μm) and typical acquisition periods of 200–500 ms. The Lumencor light sources produce little ultraviolet or infrared light, obviating the need for supplementary filters to block these wavelengths. Cells were allowed to equilibrate in the device for several hours before imaging, and all data before the first chain or pulse in each lineage was ignored in subsequent analysis. Images were acquired every 10 min and saved as 16 bit TIFFs. Focal drift was controlled through the use of the Nikon PerfectFocus system and a custom-built, image-based autofocus that imaged a sacrificial position over many planes.

Induction of chaining with IPTG. To induce chaining, two syringes carrying either LB with 0.1 mg ml^{-1} BSA or LB with 0.1 mg ml^{-1} BSA and 100 μM IPTG (isopropyl β-D-1-thiogalactopyranoside) were connected via soft tubing to a Y-junction connector that fed into a common line connected to the device. The line that was not in use was clamped shut with a binder clip. Each syringe was loaded into an independently controlled syringe pump, and a pulse of IPTG was produced by switching to the IPTG-bearing syringe for 10 min.

Image processing and lineage construction. All data analysis was based on a custom Matlab image processing pipeline described in detail in the Supplementary Information. For each image, the top cell in each channel was identified as summarized in Extended Data Fig. 10. The mean fluorescence intensity within these cells was then calculated for each fluorescence channel. A simple tracking algorithm was used to follow cells as they grew and divided, producing long lineages lasting the duration of the experiment. Cell division events were identified by looking for instances where a cell's calculated area dropped to less than 60% of its previous value. If a tracked cell died spontaneously, the algorithm continued the lineage from the dead cell's closest relative.

Measuring residence times in the two states. Motility and chaining durations were called by examining the trace of P$_{tapA}$-*cfp* fluorescence within a lineage. To identify the level of background fluorescence, rough peaks were identified using a peak-finding algorithm (N. C. Yoder, available at http://www.mathworks.com/matlabcentral/fileexchange/25500-peakfinder) on traces smoothed with a Savitzky–Golay filter, and the average fluorescence outside these peaks was subtracted from all traces. Final peak boundaries were called where the matrix reporter signal crossed pre-defined thresholds. These thresholds were chosen to correspond to phenotypic transitions: onset of matrix gene expression defines the beginning of the peak, and onset of motility gene expression defines the end (Extended Data Fig. 1). We note that the main conclusions of the paper are insensitive to the threshold values (Extended Data Fig. 1). All peaks were manually curated before calculating statistics.

With the cell-fate history of each lineage in hand, we compiled statistics describing residence time in the chained state (chain/pulse periods) and residence time in the motile state (subsequent initiation times and motility periods). We define a chain or pulse period as the duration of matrix expression within a peak (identified as described above) and the motility period as the duration of uninterrupted motility gene expression between chaining events. In Fig. 3b, we instead measured the time between the start times of consecutive peaks ('subsequent initiations'), meaning either chains or pulses. Owing to the long average residence time in the motile state, long motility periods are difficult to sample adequately. We account for this issue in the calculation of motility-related statistics, and include a complete discussion of the correction in the Supplementary Information.

Log transformation. We define the log transformation of a cumulative distribution function $F(t)$ as $-\log[1 - F(t)]$. This transformation facilitates comparisons, as exponential distributions are transformed to straight lines.

Memory (mean residual lifetime). We measured the memory associated with each state using the mean residual lifetime, defined as $m(t) = E[T - t \,|\, T > t]$ for a distribution of residence times, T. The mean residual lifetime at time t is the average amount of time a cell will remain in its current state given that it has already spent t time units there.

Average expression profiles. Average profiles of matrix gene expression during pulses and chains were created by normalizing all measured events' heights to 1, aligning the events' leading edges, and averaging the expression values at each time point. This procedure normalizes away variability in peak height so that variation between average traces derives primarily from differences in timing.

Identifying chain build-up and dilution phases. Each chaining event was decomposed into 'build-up' and 'dilution' phases based on rates of matrix reporter synthesis and dilution that were calculated from each trace. Briefly, traces were smoothed using a Savitzky–Golay filter, the resulting polynomial was differentiated, and the rate of expression was inferred from a kinetic model of gene expression (see Supplementary Information) that assumed a time varying synthesis rate and exponential degradation of reporter. The build-up phase was defined as the time over which the synthesis rate of reporter was at least 20% of the dilution rate, and the dilution phase was the remaining time in which dilution dominated.

51. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using microManager. *Curr. Protoc. Mol. Biol.* Ch. 14, Unit14.20 (2010).

a

b
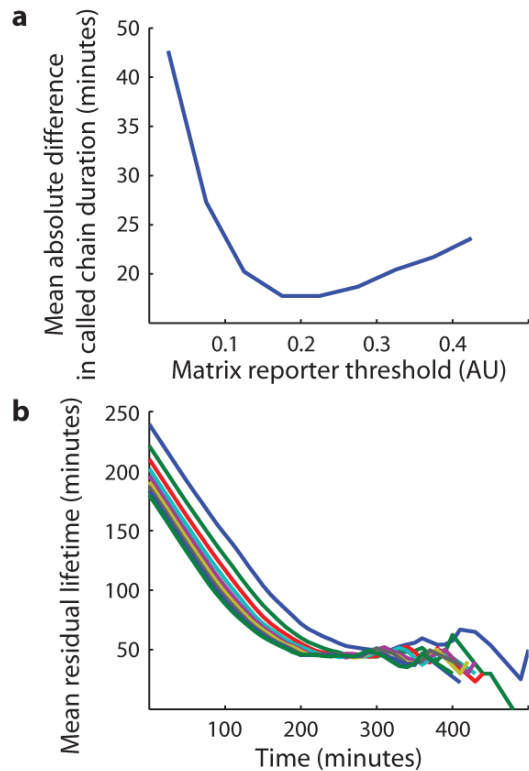
Extended Data Figure 1 | Ageing behaviour is independent of choice of threshold. Initially, the duration of a chaining event was called as the time between when matrix expression was first detectable to when flagellin expression began to increase. However, to compare chains (in strain TMN1157) and pulses (in strain TMN1158), we examined whether it was possible to call the end point using only the matrix reporter, as flagellin expression does not fall during pulses. In both methods, the beginning of a chain was called as the time when the matrix signal was first detectable above background fluctuations ($\sim$0.033 arbitrary fluorescence units, AU; see Supplementary Information). **a**, To call the end of a chain using only the matrix signal, various thresholds were applied. The figure plots the difference in chain duration between this single reporter method for different thresholds and the two reporter method. A threshold of 0.15 AU called the duration of chaining to within 20 min of the two-reporter method and was used throughout the text to call the end of the events. **b**, To show that the primary conclusions are unchanged by the choice of threshold, the ageing curves for the chained state are plotted for all thresholds shown in the previous panel. As the motile state is extremely long in comparison to the chained state, properties of the motile state are completely insensitive to how we called chains.
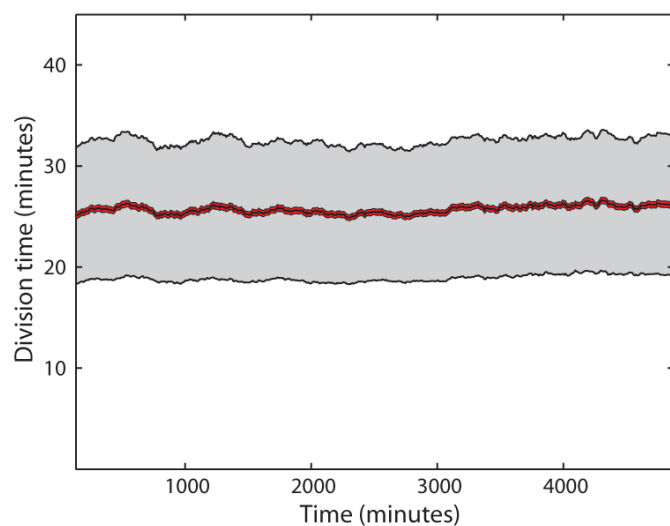
Extended Data Figure 2 | Cell growth is homogeneous in time. Sliding window average of division time plotted as a function of time (in strain TMN1158). Each point in the curve represents the average of all division times that occurred within a 250-min window. The grey shaded area denotes ±1 standard deviation, whereas the red shaded error denotes ±1 standard error of the mean. A flat trend indicates that conditions in the device do not change in time.
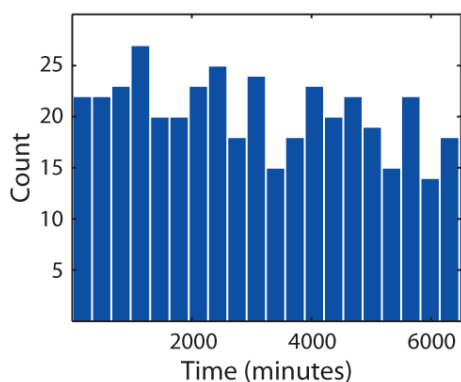
Extended Data Figure 3 | Chaining incidence is constant in time. Histogram of the number of chaining events observed in successive 330-min windows in the experiment described in Fig. 2 of the main text. As the number of observed lineages was constant throughout the experiment, these measurements reflect the average chaining rate in each window. A flat trend occurs when this average rate is constant in time, and thus that the factors controlling the switching decision have reached stationarity. Chains occurring early in our experiments were excluded from subsequent analysis to avoid any transient effects associated with adapting to growth in the device (Supplementary Information).
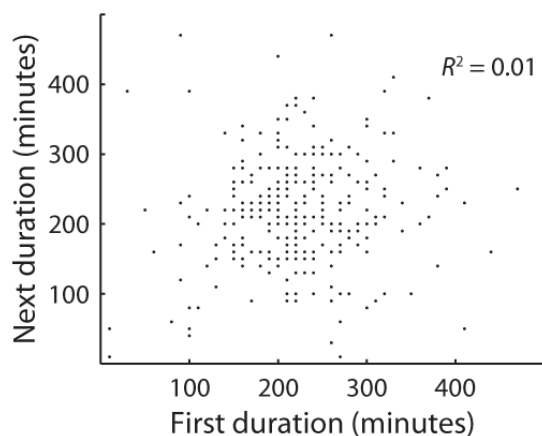
Extended Data Figure 4 | Successive visits to the chained state are uncorrelated. Scatter plot of the durations of sequential visits to the chained state within each wild-type lineage (440 events), analogous to Fig. 2b for the motile state. Note that some points fall on top of each other owing to the discrete nature of the measurements.
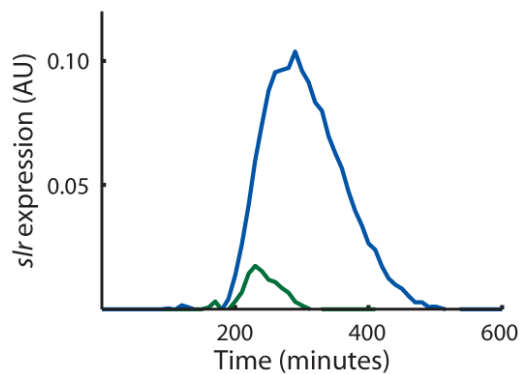
Extended Data Figure 5 | SlrR is expressed strongly only in chains. Average expression traces of *slrR* during chains (blue curve, 25 events) and pulses (green curve, 14 events) seen in strain TMN1180 (P*tapA*-*cfp* P*slrR-mKate2 hagA233V*). AU, arbitrary units.
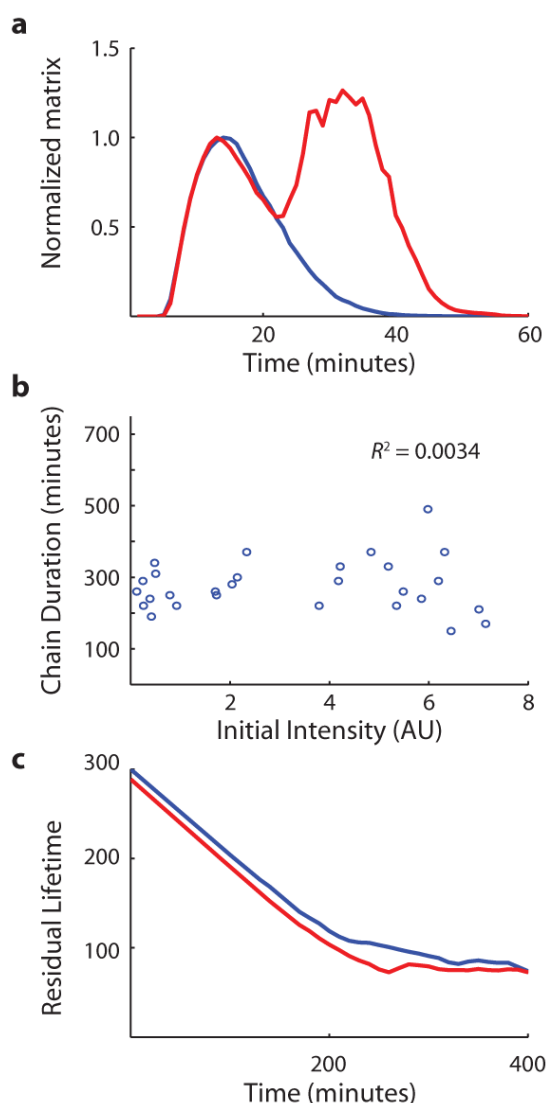
Extended Data Figure 6 | Chaining program is independent of cellular state.
To test whether the initial state of the cell influenced the chaining program, cells (of strain TMN1195) were forced to chain with a burst of expression from an IPTG-inducible *sinI* gene (created by switching to medium containing 100 μM IPTG for 10 min). When some cells began to return to the motile state (3 h later), a second IPTG treatment was administered. **a**, Average matrix expression profiles in chains induced by single pulses of IPTG (blue curve) or two consecutive IPTG pulses (red curve). The average amount of time spent as a chain after the second IPTG treatment was similar to the time seen in the chained state after a single treatment (260 min versus 280 min, 177 and 28 events, respectively). **b**, Scatter plot comparing matrix expression level (in arbitrary fluorescence units, AU) at the time of the second IPTG treatment to the duration of the ensuing chain, indicating that the state of the cell at the time of treatment had no effect on the subsequent chain duration. **c**, 10 min (blue curve, 84 events) and 20 min (red curve, 99 events) IPTG treatments were used to induce chaining, resulting in near identical distributions of chain durations. Note that the 10-min data set contained two exceptionally long chaining events that explain the slightly higher average duration.
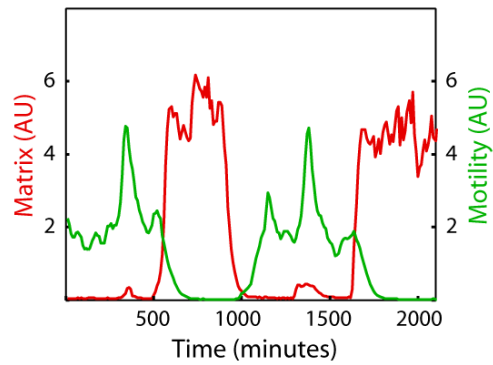
Extended Data Figure 7 | Strongly enhanced commitment to the chained state in strains overexpressing *slrR*. The figure shows an example trace of a chain made by the strain TMN1206 (P_{tapA}-*cfp* P_{hag}-*mKate2 hagA233V ywrK*::P_{slrR}-*slrR*), which bears an additional copy of the gene for SlrR under its native promoter. In this strain, most chains last long enough that they are eventually pulled out by the flow of fresh medium running through the device. Using the time to fall-out as a lower bound for the average duration of the chaining state suggests that the chained state lasts at least ~420 min (~15.5 generations) in these cells. AU, arbitrary units.
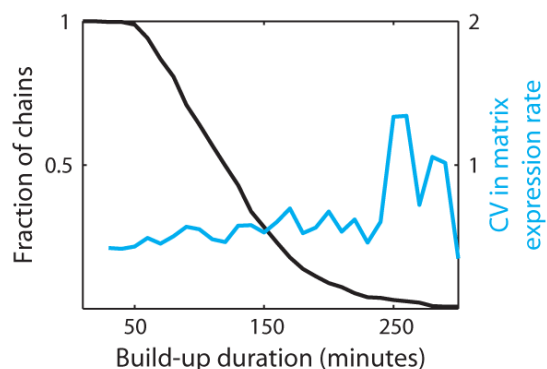
Extended Data Figure 8 | Variation in matrix expression rate over time during build-up phase. As described in the main text, chaining events can be naturally broken down into a build-up period, when new synthesis dominates, and a subsequent dilution period, where new synthesis is minimal. The rate of matrix reporter expression was calculated at each time point during the build-up period for all chaining events, producing a time-varying distribution of possible expression rates. The figure plots the coefficient of variation of this distribution, showing that expression rates show a roughly constant CV of ~0.5 over much of the build-up period. Note that most chains have ceased the build-up phase by about 250 min in, so the end of the graph is less informative. This figure should be compared with Fig. 4f, which shows that the CV in the abundance of the matrix reporter decreases over the same period due to the time averaging principle described in the main text.
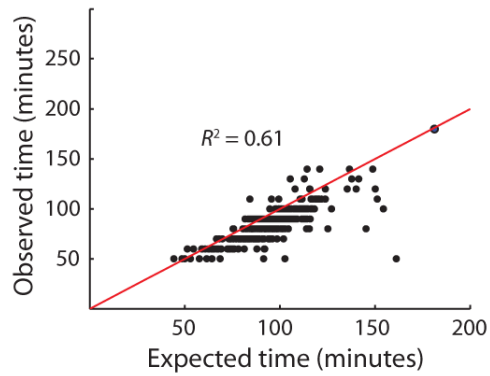
Extended Data Figure 9 | Dilution phase is well described by a deterministic model for exponential decay. Scatter plot comparison of observed and predicted dilution phase durations in spontaneous chains. Expected dilution times were derived from a deterministic model for exponential decay of the reporter (Supplementary Information). Close proximity to the line $y = x$ (red line) indicates that the data are well described by the model.
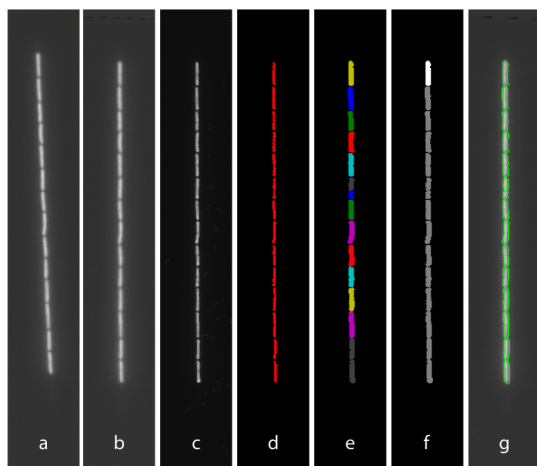
Extended Data Figure 10 | Image processing used for image quantification.
a, Cells are identified using a constitutively expressed YFP construct. **b**, Images are rotated so that channels are oriented vertically. **c**, Images are contrast enhanced to better identify cell boundaries. **d**, Cells are preliminarily identified by edge detection. **e**, The mask identifying cells is improved by morphological processing. **f**, Mother cells are identified (highlighted in white). **g**, Superposition of segmented cell boundaries and rotated data YFP image.

Effect of natural genetic variation on enhancer selection and function

S. Heinz[1]*, C. E. Romanoski[1]*, C. Benner[1,2,3], K. A. Allison[1], M. U. Kaikkonen[1,4], L. D. Orozco[5] & C. K. Glass[1,3,6]

The mechanisms by which genetic variation affects transcription regulation and phenotypes at the nucleotide level are incompletely understood. Here we use natural genetic variation as an *in vivo* mutagenesis screen to assess the genome-wide effects of sequence variation on lineage-determining and signal-specific transcription factor binding, epigenomics and transcriptional outcomes in primary macrophages from different mouse strains. We find substantial genetic evidence to support the concept that lineage-determining transcription factors define epigenetic and transcriptomic states by selecting enhancer-like regions in the genome in a collaborative fashion and facilitating binding of signal-dependent factors. This hierarchical model of transcription factor function suggests that limited sets of genomic data for lineage-determining transcription factors and informative histone modifications can be used for the prioritization of disease-associated regulatory variants.

Inter-individual genetic variation is a major cause of diversity in phenotypes and disease susceptibility. Although sequence variants in gene promoters and protein-coding regions provide obvious prioritization of disease-causing variants, most (88%) genome-wide association study (GWAS) loci are in non-coding DNA, suggesting regulatory functions[1]. Prioritization of functional intergenic variants remains challenging, owing in part to an incomplete understanding of how regulation is achieved at the nucleotide level in different cell types and environmental contexts[2–11]. Recent studies have described important roles for lineage-determining transcription factors (LDTFs), also referred to as pioneer factors or master regulators, in selecting cell-type-specific enhancers[12–15], but the sequence determinants that guide their binding are poorly understood. Previous findings in macrophages and B cells suggest a hierarchical model of regulatory function[6], in which a relatively small set of LDTFs collaboratively compete with nucleosomes to bind DNA in a cell-type-specific manner (Fig. 1A, a and b). The binding of these factors is proposed to 'prime' DNA by initiating deposition of histone modifications that are associated with *cis*-active regulatory regions (Fig. 1A, b and c) and enable concurrent or subsequent binding of signal-dependent transcription factors that direct regulated gene expression[6,13,15,16] (Fig. 1A, c–e). In principle, this model provides a straightforward framework that allows non-coding variants to be classified with respect to their ability to directly perturb LDTF binding and their potential to exert indirect effects on binding of other LDTFs and signal-dependent transcription factors. To test the validity of this model and its ability to explain effects of genetic variation on transcription factor binding and function, we exploited the naturally occurring genetic variation between the inbred C57BL/6J and BALB/cJ mouse strains (~4 million single nucleotide polymorphisms (SNPs) and ~750 k indels[17]) as an '*in vivo* mutagenesis screen'.

Direct effects of genetic variation

First, we quantified genome-wide binding patterns of macrophage LDTFs PU.1 and C/EBPα from both mouse strains using chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq). These experiments identified a combined 82,154 PU.1 and 54,874 C/EBPα peaks, with less than 1% of sites exhibiting highly significant strain-specific binding (PU.1, $n = 496$; C/EBPα, $n = 263$; fourfold tag count ratio, false discovery rate (FDR) $< 1 \times 10^{-14}$, >90% located >3 kilobases (kb) from gene promoters) (Fig. 1B, C and Extended Data Fig. 1a). Strain-specific binding was defined using biological ChIP-Seq replicates, which yielded <0.2% empirical false positives (Extended Data Fig. 1b–g). Differential binding of PU.1 and C/EBPα was significantly correlated with differential expression of the nearest gene as measured by RNA-Seq (Fig. 1D). There were no apparent differences in genomic context for strain-similar and strain-specific binding at inter- or intragenic sites (>3 kb to promoters) as defined by CpG content, distance from nearest gene or repetitive element, or conservation score (Extended Data Fig. 2a). Instead, strain-specific binding was highly correlated with polymorphism frequency. We observed fivefold enrichment of polymorphisms at strain-specific versus strain-similar PU.1- and C/EBPα-bound regions (Fig. 1E and Extended Data Fig. 2b), with the greatest variant density at the peak centres (Extended Data Fig. 2c, d).

To investigate the direct effects of sequence variants on transcription factor binding, we identified the most enriched position weight matrices (PWMs) in genomic regions marked by histone H3 lysine 4 di-methylation (H3K4me2) or bound by PU.1 or C/EBPα (Extended Data Fig. 3a and Supplementary Table 1). This analysis consistently identified consensus and degenerate motifs for the LDTFs PU.1, C/EBP and AP-1 as the most highly enriched PWMs. Notably, the frequency of mutations in these motifs increased with strain-specific binding of PU.1 and C/EBPα (Extended Data Fig. 2e, f). Excluding strain-specific loci without *cis*-variation (~11%), 41% of strain-specific PU.1 binding directly associated with strain-specific mutations in PU.1 motifs in the other strain. For C/EBPα, 44% of strain-specific binding associated with strain-specific C/EBPα motifs (Fig. 1F).

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0651, La Jolla, California 92093, USA. [2]Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [3]San Diego Center for Systems Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. [4]Department of Biotechnology and Molecular Medicine, A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, PO Box 1627, 70211 Kuopio, Finland. [5]Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, 3000 Terasaki Life Sciences Building, Los Angeles, California 90095, USA. [6]Department of Medicine, University of California, San Diego, 9500 Gilman Drive, Mail Code 0651, La Jolla, California 92093, USA.
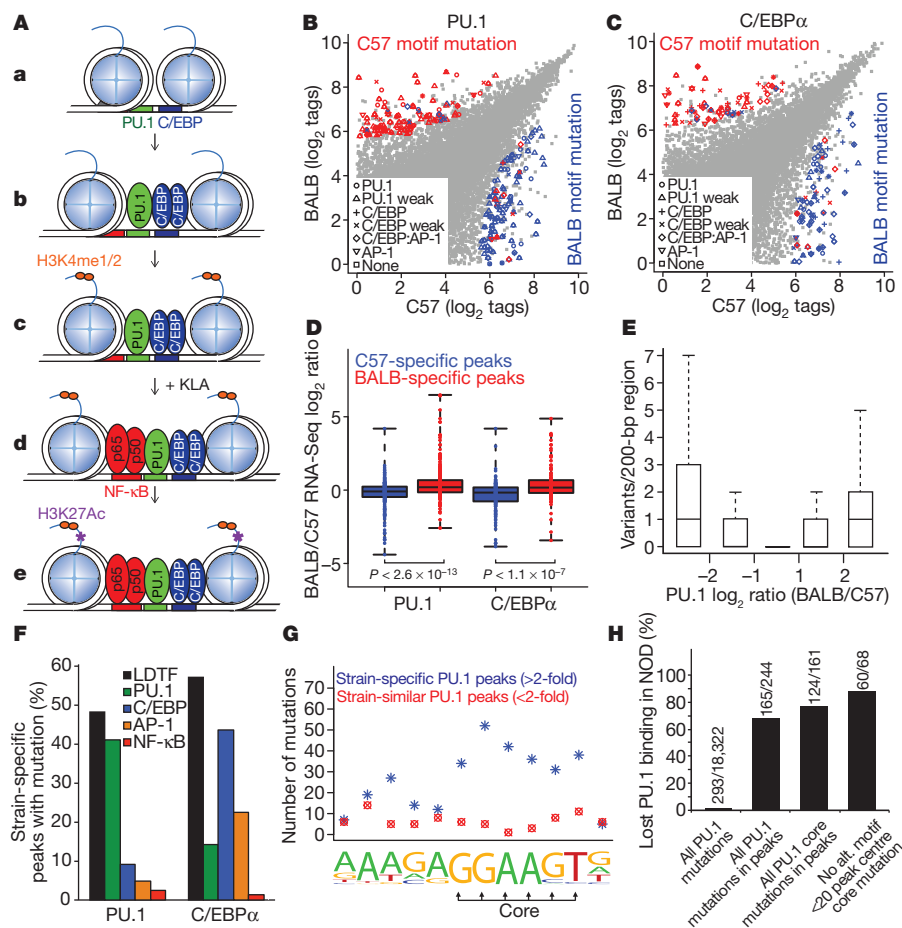*These authors contributed equally to this work.

Figure 1 | Genetic variation affects LDTF binding. **A**, Model in which LDTFs (PU.1 and C/EBPα) establish regulatory function (explained in text). **B**, **C**, ChIP-Seq-defined binding intensity for PU.1 (**B**) and C/EBPα (**C**) in resting macrophages derived from C57BL/6J (C57; *x* axes) and BALB/cJ (BALB; *y* axes) mice. Dots represent normalized tag counts in 200-bp peaks. PU.1, C/EBPα and AP-1 motifs that were mutated in one genome (distinguished by symbol; red denotes C57BL/6J, blue denotes BALB/cJ) are highlighted for peaks with strain-specific binding (fourfold, FDR $\leq 1 \times 10^{-14}$). **D**, RNA-Seq-determined expression for genes nearest to strain-specific PU.1 or C/EBPα peaks. *P* values are from one-tailed *t*-test. **E**, Variant frequency distributions for PU.1 binding ratio bins. Box midlines (**D**, **E**) are medians, boundaries are first and third quartiles, and whiskers extend to extremes. **F**, The percentage of polymorphic, strain-specific PU.1 and C/EBPα peaks with LDTF mutations. **G**, The observed position of SNPs generating strain-specific PU.1 motifs (*n* = 359) underlying differential (blue) or similar (red) PU.1 binding are shown. **H**, The proportion of NOD PU.1 motif mutations that abolished PU.1 binding for each group is shown (details in Extended Data Fig. 5). Alt., alternative.

Although strain-specific binding of PU.1 and C/EBPα was highly linked to strain-specific motif mutations, strain-specific motif mutations were also associated with strain-similar binding (Extended Data Fig. 3c, d). This raised the question as to whether specific features of motif mutations could be used to predict strain-specific binding. Comparison of motif mutations in strain-specific and strain-similar peaks revealed three distinct attributes contributing to predictive power. First, mutated motifs within 20 base pairs (bp) of the experimentally defined binding centres were more highly associated with an effect on binding (PU.1, $P = 1.6 \times 10^{-4}$; C/EBPα, $P = 0.036$; Extended Data Fig. 4a–d). Second, the presence of alternative motifs within 100 bp of the PU.1 peak centres significantly buffered the effect of strain-specific PU.1 motifs (Extended Data Fig. 4e, f). Third, after removing peaks with alternative motifs, analysis of the nucleotides mutated enabled delineation of an empirically defined functional motif that revealed a strong relationship between 'core' mutations and altered binding (Fig. 1G and Extended Data Fig. 4g–i; $P = 3.2 \times 10^{-8}$ PU.1 and $P = 5.1 \times 10^{-4}$ C/EBP). Taken together, core motif mutations <20 bp from the peak centre that lacked alternative motifs were 3.5× and 3× more likely to occur in differential versus similar bound peaks for PU.1 and C/EBPα, respectively (Extended Data Fig. 4j, k). Notably, up to 90% of these mutations were located in differentially bound peaks (Extended Data Fig. 4l, m). To investigate the possibility that an algorithm incorporating these characteristics could be used to predict the effect of a specific motif mutation on transcription factor binding, we performed ChIP-Seq analysis for PU.1 in macrophages derived from a third inbred strain of mice, NOD/ShiLtJ (NOD). Of the ~1.4 million identifiable PU.1 motifs in the C57BL/6J reference genome, 18,322 contain SNPs that mutate the PU.1 motif in the NOD genome. A total of 1.6% of these mutations were associated with strain-specific binding (Fig. 1H). Of the 244 NOD PU.1 motif mutations

located in PU.1-bound regions in C57BL/6J or BALB/cJ mice, 68% were associated with strain-specific binding. When considering all three variables (motif distance, alternative motif and motif core; Extended Data Fig. 5), 88% of the predicted functional mutations were consistent with impaired PU.1 binding in NOD (Fig. 1H).

Variation and collaborative LDTF binding

To investigate the potential effect of mutations in LDTF recognition motifs on collaborative binding, we analysed all strain-specific PU.1 or C/EBPα binding events in regions containing LDTF motif mutations. PU.1 motif mutations resulting in loss of PU.1 binding were frequently associated with a corresponding loss of nearby C/EBPα binding in the absence of C/EBP motif mutations (Fig. 2a, top). Conversely, C/EBP motif mutations resulting in a loss of C/EBPα binding were frequently associated with a corresponding loss of nearby PU.1 binding in the absence of PU.1 motif mutations (Fig. 2a, middle). Similar results were observed at locations containing strain-specific mutations in AP-1 binding motifs, but intact PU.1 and C/EBP motifs (Fig. 2a, bottom).

We next considered the global relationships of mutations in PU.1, C/EBP and AP-1 motifs with strain-specific binding of PU.1 and C/EBPα, taking into account both consensus and 'weak' motifs for PU.1 and C/EBP. NF-κB motifs were included as controls that were not expected to affect PU.1 or C/EBPα binding in unstimulated macrophages (Extended Data Fig. 3a, b and Supplementary Table 1). Although mutations in PU.1 motifs had the strongest effect on strain-specific PU.1 binding, mutations exclusively in C/EBP and/or AP-1 motifs also significantly correlated with differential PU.1 binding relative to similarly bound loci (Fig. 2b). Similar relationships were observed for C/EBP (Extended Data Fig. 6a). The motif distance distributions for co-bound factors were broad (half-width ~100 nucleotides), and only a minor subset of sites exhibited defined distances expected for direct protein–protein
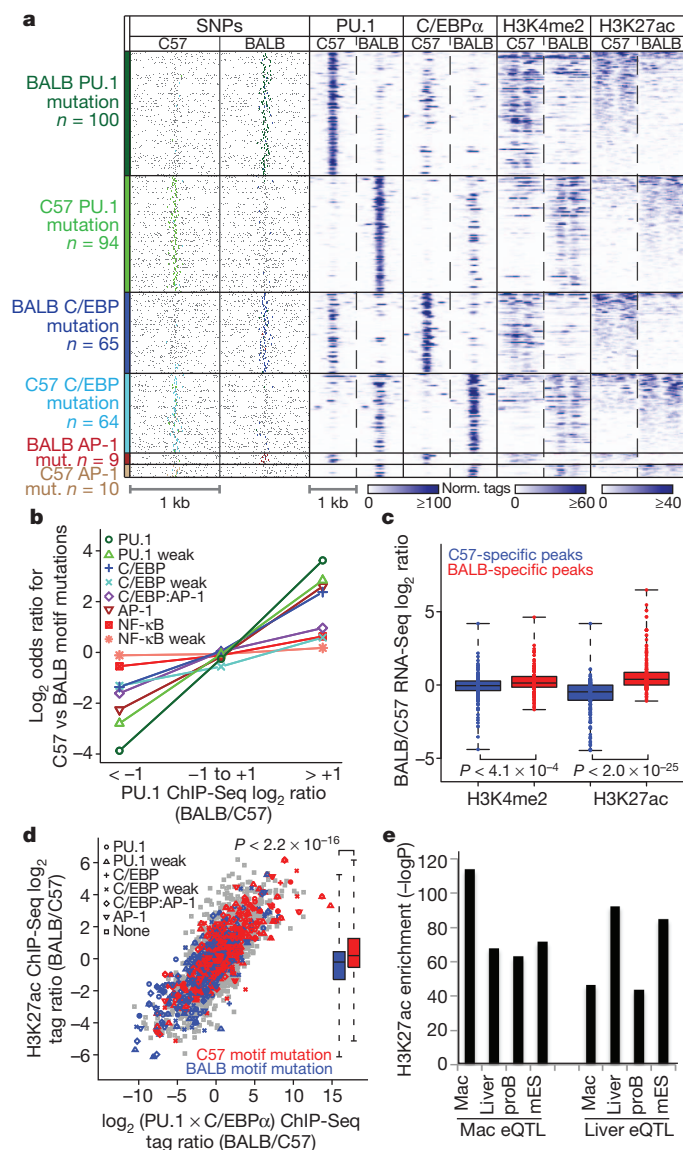
motif mutation (Fig. 1F). To test genetically whether these correlations are consistent with a collaborative binding model, we considered all LDTF motif mutations and evaluated their effects on PU.1 binding in macrophages derived from NOD mice. For polymorphic strain-specific PU.1 loci containing strain-specific LDTF motifs ($n = 220$), PU.1 binding profiles matched the strain with shared alleles for 91% of cases (Fig. 3a). At 8% ($n = 17$) of the loci, the NOD genome broke the C57BL/6J–BALB/cJ haplotypes, and in all cases, the NOD genotype at the LDTF motif variant matched the strain with similar binding (Supplementary Table 2), indicating that these variants are probably the cause of binding differences. An example is shown in Fig. 3b, in which PU.1 binds in C57BL/6J but not in BALB/cJ or NOD mice. Only one SNP in this region is associated with PU.1 binding exclusively in C57BL/6J; here, the T allele forms part of a neighbouring AP-1 motif in C57BL/6J that is mutated by the C allele in BALB/cJ and NOD mice. These findings provide genetic evidence that PU.1 binding to this location is dependent on collaborative interactions with AP-1.

To confirm that the allele-specific binding also occurs in heterozygous cells, we performed ChIP-Seq for PU.1 and C/EBPα in macrophages from CB6F1/J hybrid mice, which are F_1 offspring of a C57/BL/6J × BALB/cJ cross. In the most cases, alleles bound specifically in a parental strain were also bound preferentially in the F_1 generation (Fig. 3c and Extended Data Fig. 6c).

Figure 2 | Genetic variation supports the LDTF collaborative binding model. a, Normalized ChIP-Seq signal at 342 loci defined by strain-specific PU.1 and/or C/EBPα binding, and containing LDTF motif mutations (rows) plotted for each factor/modification (columns). Left columns display SNPs as grey dots with mutated motifs highlighted by colour (LDTF mutation labels at left). **b**, Log₂ odds ratios for observing strain-specific motif mutations at strain-specific (>2-fold tag ratio, left and right bins) and similar (<2-fold tag ratio, middle bin) PU.1 peaks (details in Methods). **c**, Gene expression for genes nearest promoter-distal (>3 kb), strain-specific H3K4me2 and H3K27ac peaks is shown (described in Fig. 1D). **d**, Normalized H3K27ac log₂ tag ratios (1 kb, y axis) versus log₂(PU.1 × C/EBPα) tag-strain ratios (200 bp, x axis) for loci with PU.1 or C/EBPα binding. Strain-specific motif mutations are indicated by symbol and colour. The distribution of H3K27ac strain ratios stratified by strain mutations is shown (two-sided t-test). **e**, Enrichment significance (hypergeometric distribution testing, see Methods) of H3K27ac-modification in eQTLs from different cell types is shown. Mac, macrophage.

interactions (Extended Data Fig. 6b), suggesting transcription-factor–nucleosome competition as the driving force behind the collaborative binding behaviour[6,18]. Together, strain-specific mutations in nearby C/EBP and AP-1 motifs were associated with ~15% of strain-specific PU.1 binding at sites with strain-similar PU.1 motifs. Mutations in nearby PU.1 and AP-1 motifs were associated with ~30% of strain-specific C/EBPα binding at sites with strain-similar C/EBP motifs (Fig. 1F). Overall, 48% of strain-specific PU.1 binding and 57% of C/EBPα binding were associated with at least one assignable LDTF

Figure 3 | Validation of predicted binding and modification patterns. a, PU.1 binding at strain-specific loci is shown for C57BL/6J, BALB/cJ and NOD/ShiLtJ (NOD) mouse macrophages (columns; red denotes binding, white denotes no binding). All loci contain a strain-specific PU.1, C/EBP or AP-1 motif. The NOD haplotype at these loci is indicated by the sidebar (red denotes BALB/cJ, blue denotes C57BL/6J, yellow denotes a mixture). **b**, PU.1 binding, SNPs (lines), allele sharing, motif alignment and genome sequence (seq.) are shown for a locus where NOD broke the C57/BALB haplotypes. **c, d**, Allele-specific ChIP-Seq ratios (y axes) for PU.1 (**c**) and H3K27ac (**d**) in CB6F1/J hybrid macrophages versus ChIP-Seq reads in parental strains (BALB/C57 log₂ ratios; x axes) are shown for strain-specific (sp.) peaks (blue denotes C57BL/6J, red denotes BALB/cJ-specific) and similarly (sim.) bound peaks (black) as defined by parental data.
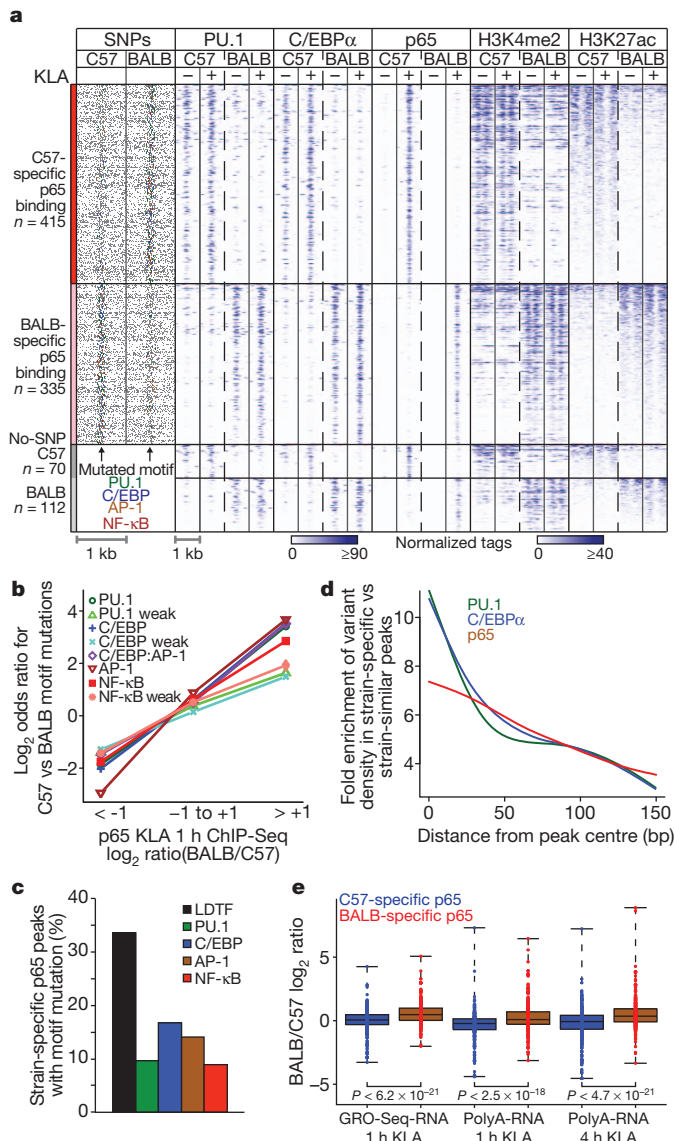
Figure 4 | p65 binding is largely determined by LDTF binding. a, Strain-specific p65-bound regions were segregated into rows according to the bound strain (coloured side bar). Binding/modification is shown with and without $100\ \mathrm{ng\ ml^{-1}}$ KLA treatment ($+/-$, third header row). As in Fig. 2a, SNPs are indicated by grey dots and mutated motifs are highlighted by colour (labelled at bottom). **b**, The $\log_2$ odds ratio for observing strain-specific mutations is shown for bins of p65 binding as described in Fig. 2b. **c**, The percentage of polymorphic, differentially bound p65 loci containing LDTF or NF-κB motif mutations is shown. **d**, The ratio of variant counts in strain-specific versus strain-similar peaks (y axis) is shown relative to the peak centres for PU.1-, C/EBPα- and p65-bound peaks in 10-bp bins (x axis), smoothed using cubic spline. **e**, The relative amount of transcription (GRO-Seq) and mRNA production between strains after KLA treatment at the nearest gene to strain-specific p65 loci is shown. P values are from one-tailed t-test.

Given the genetic evidence that LDTFs collaborate to bind DNA, we next tested the extent to which strain-specific LDTF binding explained promoter-distal (>3 kb) strain-specific histone modification events, such as H3K4me2 and H3K27ac deposition, which respectively mark 'primed' and 'active' chromatin[19,20] (Fig. 1A, b and c). Genomic regions exhibiting strain-specific binding of PU.1 and C/EBPα were associated with strain-specific H3K4me2 and H3K27ac (Fig. 2a, right). Strain-specific histone modifications correlated with nearby gene expression (Fig. 2c), and H3K27ac modification tracked with the corresponding parental allele in CB6F1/J hybrid mice (Fig. 3d). Strain-specific binding of PU.1 and C/EBPα was individually correlated with H3Kme2 and

H3K27ac deposition, with the combined binding of both factors exhibiting even greater correlation than the individual factors (Extended Data Fig. 7a–f). Furthermore, LDTF motif mutations segregated with differential LDTF binding and histone modifications (Fig. 2d and Extended Data Fig. 7g). Together, these findings support the concept that LDTFs have quantitatively important roles in establishing these histone modifications, probably through initiating transcription in a combinatorial fashion[21].

Expression quantitative trait loci (eQTLs) are polymorphic loci whose alleles are associated with individual RNA expression levels across a population[22]. Thus, eQTLs define active gene regulatory loci and provide an alternative method for assigning regulatory function to gene expression. To interrogate the relationship between histone modification and eQTLs, we analysed previously reported eQTL data from 85 inbred mouse strains in the hybrid mouse diversity panel in primary macrophages[23] (see Methods). We found that eQTLs overlapped H3K4me2- or H3K27ac-marked regions at frequencies greater than expected by chance, supporting the role of histone modifications as landmarks of regulatory activity (hypergeometric test P values: H3K4me2 $= 1 \times 10^{-2,147}$, H3K27ac $= 1 \times 10^{-2,290}$. Next, given the highly cell-type-specific nature of gene regulation[24], we proposed that eQTLs from different cell types would be reflected in the histone modification profiles in the same cell type. We examined liver and macrophage eQTLs for a set of ~130 k SNPs from the hybrid mouse diversity panel[25] for overlap with H3K27ac loci defined in macrophages or in liver, pro-B or mouse embryonic stem cells[20]. Macrophage eQTLs were more significantly enriched for overlap with macrophage H3K27ac regions than liver H3K27ac regions. Similarly, liver eQTLs were most significantly enriched with liver H3K27ac relative to macrophage H3K27ac (Fig. 2e). Clustering of H3K27ac profiles revealed that liver and embryonic stem-cell H3K27ac profiles are most similar (Extended Data Fig. 7h), providing an explanation as to why liver eQTLs were highly enriched in mouse embryonic stem-cell H3K27ac regions.

LDTF motif mutations affect NF–κB binding

To evaluate the prediction that primed regulatory loci (containing H3K4me2) often require additional binding of signal-dependent transcription factors to achieve regulatory activity (Fig. 1A, c–e), we treated C57BL/6J and BALB/cJ macrophages with Kdo$_2$-lipid A (KLA), a potent and specific agonist of TLR4 (ref. 26). KLA treatment causes NF-κB to enter the nucleus, bind DNA and regulate several hundred target genes[26,27]. We performed ChIP-Seq for PU.1, C/EBPα and the p65 (also known as RelA) component of NF-κB in untreated and KLA-treated macrophages, and observed that 61% of sites that gained p65 were pre-bound by PU.1 and/or C/EBPα without KLA. *De novo* motif analysis indicated that an AP-1 motif was present in 42% of the remaining sites, suggesting that AP-1 is responsible for priming a large proportion of the p65 cistrome (Extended Data Fig. 8a), in line with previous reports[16].

To interrogate the dependence of p65 on LDTFs further we focused on sites that gained p65 only in one strain ($n = 932$, >90% promoter-distal; Extended Data Fig. 1a, Fig. 4a, fourth column). In most cases, PU.1 and/or C/EBPα were bound before KLA treatment only in the strain exhibiting p65 binding (Fig. 4a). In addition, strain-specific p65 binding primarily occurred at loci already marked by H3K4me2, and led to an increase of H3K27ac, consistent with the proposed model. To analyse the effects of genetic variation on transcription factor motifs, we performed strain-specific LDTF and NF-κB motif finding in polymorphic strain-specific p65-bound peaks ($n = 750$) (Extended Data Fig. 3b). Notably, p65 binding was influenced by mutations in individual LDTF motifs to a similar extent as mutations in the NF-κB motif itself (Fig. 4b and Extended Data Fig. 8b). For strain-specific p65 binding events, 34% could be attributed to assignable mutations in PU.1, C/EBP or AP-1 motifs, whereas 9% could be explained by mutations in the assignable NF-κB motifs themselves (Fig. 4c). RelA is known to bind to degenerate and non-canonical motifs[28] that might

not be captured by *de novo* motif analysis. To gain motif-independent insight into variant location and strain-specific transcription factor binding, we assessed the variant frequency relative to the centres of strain-specific p65 peaks. Similar to strain-specific PU.1 and C/EBPα peaks, strain-specific p65 peaks are in regions of higher variant density than strain-similar peaks (Extended Data Fig. 8c). In contrast to LDTFs, in which strain-specifically bound regions have a high variant density at their peak centres, the distribution of variants at strain-specific p65 peaks is significantly different from those of the LDTFs (Kolmogorov–Smirnov $P < 0.013$), as it contains fewer variants at the peak centres and is broader (Fig. 4d and Extended Data Fig. 8d–f). This is consistent with p65 binding being more affected by sequence variation in motifs of neighbouring factors than LDTFs.

Overall, strain-specific p65-bound regulatory sites were significantly correlated with nearby genic transcription and messenger RNA production (Fig. 4e). We tested strain-specifically bound and epigenetically marked putative enhancer sequences with strain-specific mutations for differential enhancer function in transient and stable reporter assays (Fig. 5a, b and Extended Data Fig. 9a, b). We observed the predicted strain-specific enhancer activity for 18 out of 20 of these genomic sequences. Conversely, enhancer elements with sequence variation in non-core nucleotides that were not predicted to alter PU.1 or C/EBP binding and that exhibited strain-similar binding patterns exhibited similar enhancer activity (Extended Data Fig. 10a).

Lastly, we tested whether the predicted motif-disrupting variants could specifically explain strain-specific enhancer activity by swapping variants at the putative causative alleles in C57BL/6J and BALB/cJ while maintaining the genetic background for the remainder of the enhancer sequences. Representative examples in which reversal of such SNPs in PU.1, C/EBP and p65 motifs reversed strain-specific enhancer activity are illustrated in Fig. 5c and Extended Data Fig. 10b, c. By

contrast, reversal of nearby SNPs not predicted to alter LDTF motifs had no effect on strain-specific enhancer activity (Extended Data Fig. 10c).

Discussion

Together, we have exploited natural genetic variation to test a collaborative model for enhancer selection and function, and conversely explored the ability of this model to explain strain-specific differences in transcription factor binding and epigenetic features associated with functional enhancers in macrophages. These studies provide genetic evidence that LDTFs are dependent on collaborative binding to variably spaced DNA recognition motifs to select enhancers and enable binding of signal-dependent transcription factors. Notably, the variable motif distances observed at loci co-bound by LDTF suggest that collaborative binding does not generally require direct protein–protein interactions between the involved transcription factors. The proposed hierarchical LDTF collaborative model provides a conceptual framework for prioritization of non-coding disease-associated regulatory variants. Although all cells express hundreds of transcription factors, a large fraction of functional enhancers (~70% in macrophages) are characterized by collaborative interactions involving relatively small sets of lineage-determining transcription factors (for example, PU.1, AP-1 and C/EBPs). The requirement for collaborative binding interactions provides an explanation for why transcription factor binding is lost at sites where mutations do not occur in the cognate recognition motif. In the case of NF-κB, for example, mutations in the motifs for LDTFs were approximately three times more likely to result in decreased binding of NF-κB than mutations in the NF-κB-binding site itself.

An essential step in leveraging the collaborative model to pinpoint potential disease-causing variants is the definition of relevant LDTF-binding sites and functionally important variants. At the current level of genome annotation, this cannot be achieved by analysis of DNA sequence alone. For example, there are $\sim 1 \times 10^6$–2×10^6 identifiable PU.1 motifs in the human[29] and mouse genomes, but less than 10% are actually occupied by PU.1 in macrophages. By experimentally defining strain-similar and strain-specific binding patterns for PU.1, the relevant sites at which mutations can result in altered function are identified. Comparison of PU.1 motif mutations associated with strain-specific versus strain-similar binding allowed the genetic definition of a functional binding matrix and additional distinguishing features that enabled accurate prediction of functional mutations in a third strain. Thus, by collecting a relatively limited set of genomic binding data for LDTFs and informative histone modifications, this analytical approach can be exploited to explain a greater extent of variation in enhancer selection and function than previously possible[7,10]. To increase the specificity and sensitivity for detecting functional variations further, identification of transcription factor motifs that permit binding but diverge from the consensus PWM, that is, 'weak' motifs, needs to be improved, as such sites are more likely to be affected by mutation[29,30]. In addition, transcription factors less abundant than LDTFs probably have individually small but collectively significant roles. At a larger scale, non-*cis*-acting, long-range epigenetic mechanisms may also be important for enhancer selection. A major goal for the future will be to extend these approaches to understanding natural genetic variation associated with human disease.

Figure 5 | Validation of strain-specific enhancer activity and causal variants. **a**, Enhancer reporter schematic. One-kilobase enhancer-like fragments were cloned downstream of an HSV-TK-luciferase reporter gene and tested for basal and KLA-inducible transcriptional activity in RAW264.7 macrophages. **b**, Genomic features (left) and regulatory activities (right) of the strain-similar PU.1 −14-kb enhancer (enh) positive control from C57BL/6J- and BALB/cJ-derived macrophages (Extended Data Figs 9 and 10 show all 33 loci tested). Horizontal midline (left) represents the 1-kb stretch of cloned DNA, and SNPs are indicated with vertical black lines. ChIP-Seq tag pile-ups are shown for PU.1 (green), C/EBPα (blue), p65 (red), H3K27ac (purple) and H3K4me2 (orange) for C57BL/6J (above midline) and BALB/cJ (below midline) mice, with identical scales after KLA treatment (100 ng ml⁻¹, 1 h). **c**, Representative example of a strain-specific locus and the effect of a single base pair swap at the indicated PU.1 motif SNP on enhancer activity. See Extended Data Fig. 10b, c for additional examples and allele-swapping controls. Data are mean ± s.d.

METHODS SUMMARY

Cell culture. Peritoneal macrophages from male 6–8-week-old mice were thioglycolate-elicited and collected 4 days after injection[6], plated overnight, and incubated for 1 h in fresh media with or without 100 ng ml⁻¹ KLA[26].

ChIP-Seq and feature identification. ChIP-Seq was performed based on published protocols[6,16,31], on either native chromatin after MNase digestion (H3K4me2) or fixed, sonicated chromatin (PU.1, C/EBPα, H3K27ac and p65). ChIP-Seq libraries were sequenced for 51 cycles on a HiSeq 2000 sequencer (Illumina). Reads from C57BL/6J and BALB/cJ were mapped with low stringency to both the mm9 reference (C57BL/6J) genome and the BALB/cJ contigs[17], and the 98% of all reads that mapped to both genomes was kept for further analysis. NOD data were mapped to the mm9 reference with low stringency. For CB6F1/J ChIP-Seq experiments,
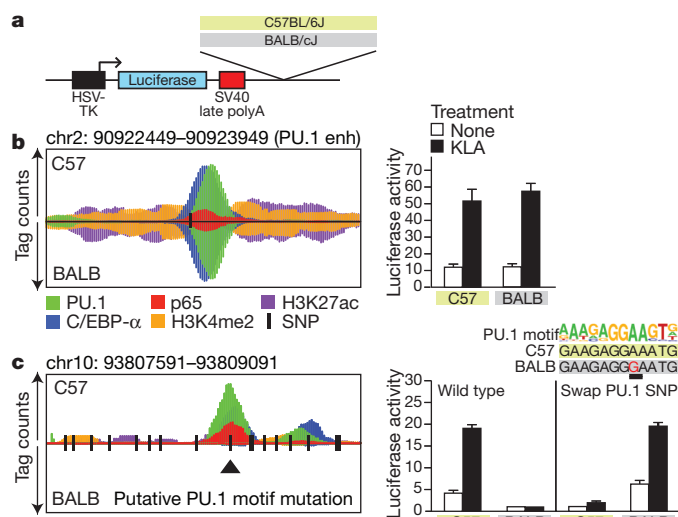
allele-specific reads were identified by alignment without mismatch exclusively to either the mm9 reference (C57BL/6J) genome or BALB/cJ contigs[17]. Data were normalized, ChIP-Seq peaks identified, motifs analysed and variants computed using the HOMER package[6] and custom R scripts.

RNA-Seq and GRO-Seq. RNA-Seq was performed on poly(A)$^+$ RNA after fragmentation, decapping, 3' and 5' adaptor ligation and reverse transcription. GRO-Seq was performed as described[32]. Data were mapped and analysed as described above.

eQTL analysis. Analysis of eQTLs was performed as previously described[23,25]. External data (GEO accession GSE24164; ref. 20) were mapped to the C57BL/6J genome as described above, and enrichment was tested using the hypergeometric distribution function.

Reporter assays. Genomic ~1-kb fragments cloned into a minimal HSV-TK promoter containing luciferase reporter plasmid were assayed in RAW264.7 macrophages after 16 h stimulation with or without 100 ng ml^{-1} KLA. Data were normalized to co-transfected UB6 promoter-driven control plasmid (transient transfections) or total protein content (stable transfections).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl Acad. Sci. USA **106**, 9362–9367 (2009).
2. Cowper-Sal·lari, R. et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nature Genet. **44**, 1191–1198 (2012).
3. Degner, J. F. et al. DNase I sensitivity QTLs are a major determinant of human expression variation. Nature **482**, 390–394 (2012).
4. Gaffney, D. J. et al. Dissecting the regulatory architecture of gene expression QTLs. Genome Biol. **13**, R7 (2012).
5. Gaulton, K. J. et al. A map of open chromatin in human pancreatic islets. Nature Genet. **42**, 255–259 (2010).
6. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell **38**, 576–589 (2010).
7. Kasowski, M. et al. Variation in transcription factor binding among humans. Science **328**, 232–235 (2010).
8. Maurano, M. T., Wang, H., Kutyavin, T. & Stamatoyannopoulos, J. A. Widespread site-dependent buffering of human regulatory polymorphism. PLoS Genet. **8**, e1002599 (2012).
9. McDaniell, R. et al. Heritable individual-specific and allele-specific chromatin signatures in humans. Science **328**, 235–239 (2010).
10. Reddy, T. E. et al. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. Genome Res. **22**, 860–869 (2012).
11. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. Genome Res. **22**, 1748–1759 (2012).
12. Garber, M. et al. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. Mol. Cell **47**, 810–822 (2012).
13. Mullen, A. C. et al. Master transcription factors determine cell-type-specific responses to TGF-β signaling. Cell **147**, 565–576 (2011).
14. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell **151**, 994–1004 (2012).
15. Trompouki, E. et al. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Cell **147**, 577–589 (2011).
16. Ghisletti, S. et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity **32**, 317–328 (2010).
17. Keane, T. M. et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature **477**, 289–294 (2011).
18. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. Proc. Natl Acad. Sci. USA **107**, 22534–22539 (2010).
19. He, H. H. et al. Nucleosome dynamics define transcriptional enhancers. Nature Genet. **42**, 343–347 (2010).
20. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl Acad. Sci. USA **107**, 21931–21936 (2010).
21. Kaikkonen, M. U. et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol. Cell **51**, 310–325 (2013).
22. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. Nature Rev. Genet. **7**, 862–872 (2006).
23. Orozco, L. D. et al. Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. Cell **151**, 658–670 (2012).
24. Song, L. et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. **21**, 1757–1767 (2011).
25. Bennett, B. J. et al. A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res. **20**, 281–290 (2010).
26. Raetz, C. R. et al. Kdo$_2$-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4. J. Lipid Res. **47**, 1097–1111 (2006).
27. Smale, S. T. Transcriptional regulation in the innate immune system. Curr. Opin. Immunol. **24**, 51–57 (2012).
28. Wong, D. et al. Extensive characterization of NF-κB binding uncovers non-canonical motifs and advances the interpretation of genetic functional traits. Genome Biol. **12**, R70 (2011).
29. Pham, T. H. et al. Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. Nucleic Acids Res. **41**, 6391–6402 (2013).
30. Jolma, A. et al. DNA-binding specificities of human transcription factors. Cell **152**, 327–339 (2013).
31. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. Cell **129**, 823–837 (2007).
32. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature **474**, 390–394 (2011).

Supplementary Information is available in the online version of the paper.

Author Contributions S.H., C.K.G. and C.E.R. designed the study; S.H., C.E.R., K.A.A., M.U.K. and L.D.O. performed experiments; C.E.R. performed all genetic-variation-related analysis; C.B. wrote custom code for HOMER2 and analysed data; K.A.A. and S.H. analysed data; C.E.R., S.H. and C.K.G. wrote the manuscript.

Author Information Data are available in the Gene Expression Omnibus (GEO) under accession GSE46494. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.K.G. (ckg@ucsd.edu).

METHODS

Animals and cell culture. Thioglycolate-elicited peritoneal macrophages were collected 4 days after injection from male 6–8-week C57BL/6J, BALB/cJ or CB6F1/J hybrid mice, and plated at 20×10^6 cells per 15-cm Petri dish in RPMI1640 plus 10% FBS and $1\times$ penicillin-streptomycin. One day after plating, cells were treated with fresh media with or without $100\,ng\,ml^{-1}$ KLA for 1 h, and then directly used for downstream analyses. All animal experiments were performed in compliance with the ethical standards set forth by University of California, San Diego's Institutional Annual Care and Use Committee (IUCAC).

ChIP-Seq and feature identification. Media were decanted and cells were fixed at room temperature with either 1% formaldehyde in PBS for 10 min (for PU.1, C/EBPα, H3K27ac ChIPs), or 2 mM disuccinimidylglutarate (DSG, Pierce) and 10% dimethylsulphoxide (DMSO) in PBS for 30 min followed by 1% formaldehyde in PBS for another 15 min (p65). After quenching the reaction by adding glycine to final concentration 0.125 M, cells were washed twice with PBS and snap-frozen in dry-ice and methanol. ChIPs for PU.1 (Santa Cruz, sc-352) and C/EBPα (Santa Cruz, sc-61) were performed as described previously[6]. The H3K27ac (Abcam, ab4729) ChIP was performed in the presence of 1 mM butyric acid. For p65 (Santa Cruz, sc-372), immunoprecipitation conditions were identical to those described before[6], except that pre-clearing was omitted, and the ChIP was performed with 5 μg antibody (Santa Cruz, sc-372) pre-bound to 50 μl Protein A Dynabeads (Invitrogen) for 30 min in 0.5% BSA in TE buffer, and a final wash with TE buffer plus 50 mM NaCl was performed before elution. ChIP-Seq library preps for the initial p65 ChIPs were performed as described previously[6]; libraries for the replicates were prepared using magnetic beads similar to described procedures[12]. ChIPs for H3K4me2 were carried out on MNase-digested chromatin as described previously[31]. To control for open chromatin and library biases, input chromatin libraries after sonication were sequenced for each strain, crosslinking condition and ChIP lysis protocol. Sequencing libraries were prepared as previously described[6] using barcoded adapters (NextFlex, Bioo Scientific), and sequenced for 51 cycles on a HiSeq 2000 sequencer (Illumina) using CASAVA1.7 or 1.8.

C57BL/6J and BALB/cJ demultiplexed fastq files were mapped to both the mm9 reference (C57BL/6J) genome and the BALB/cJ contigs[17] using Bowtie0.12.7 (ref. 33) with the options '-m 1 --best -n 3 -e 200'. Mapping parameters for C57BL/6J and BALB/cJ data allowed three mismatches in the 28-bp seed sequence, with up to five high quality mismatches in the entire read. NOD ChIP-Seq data were mapped to the mm9 genome using the above options. To identify allele-specific reads from CB6F1/J data, ChIP-seq reads were aligned to the C57BL/6J or BALB/cJ sequence using Bowtie2-2.0.0-β7 (ref. 34), allowing 0 mismatches in 32-bp reads with options '-N 0 -L 32 --score-min L,0,0 --gbar 17'. Tags mapping to both genomes were discarded. Resulting allele-specific reads were counted for regions of interest.

For C57BL/6J and BALB/cJ data, reads mapping to only one genome were discarded (<2% of total) to avoid bias caused by mapability differences, and reads mapping to both were assigned to the mm9 genomic location. Genomic binding peaks for transcription factors PU.1, C/EBPα and p65 were identified using the 'findPeaks' command in the HOMER (http://biowhat.ucsd.edu/homer/) software suite[6], with default settings of '-style factor': 200-bp peaks, with fourfold tag enrichment and 0.001 FDR significance over background (ChIP input), fourfold enrichment over local tags, and normalization to 10 million mapped tags per experiment. H3K4me2 and H3K27ac regions used for initial de novo motif finding (Extended Data Fig. 3a) were identified using the default parameters of 'findPeaks -style histone' with the addition of nucleosome-free region (nfr) centring for H3K4me2 MNase data. For H3K4me2 and H3K27ac peaks identified for comparison to LDTF binding and mutation events (for example, Fig. 2d), 'findPeaks -style peaks' was used to define more focal, non-gene-associated loci. In particular, H3K27ac regions were identified with 'findPeaks' using options '-style factor -size 1000 -L 2'(twofold enrichment over local tags). H3K27ac peaks were merged between strains using 'mergePeaks -size given', and peaks were resized to 1 kb. Peaks within 3 kb of gene promoters were excluded from further analysis. H3K4me2 peaks were identified using 'findPeaks' with options '-style factor -size 500 -L 2 -C 0' (which allows for unlimited tags considered per genome position as may occur with MNase data). Peaks were then centred on the best nfr within a 1 kb window using 'getPeakTags -nfr'. Peak files between strains were merged with 'mergePeaks -size given' and H3K4me2 tags were counted in 1-kb regions centred on the merged peak file definitions. Peaks were then re-centred based on the best nfr in 1 kb identified by 'getPeakTags -nfr' according to the strain with more H3K4me2 tags. Peaks were extended to 1 kb and restricted to those more than 3 kb from gene promoters.

Strain-specific feature and motif identification. To determine strain-specific binding and epigenetic modification events, we counted the number of sequencing tags (normalized to 10 million) at peaks/regions identified in the set of combined peaks/regions from both strains. We normalized the mean peak tag count to be equal in each strain and compared the tag counts in each region and required strain-specific peaks/regions to exhibit ≥fourfold difference in tag counts and an adjusted cumulative Poisson P value corresponding to FDR $< 1 \times 10^{-14}$ (ref. 35). These criteria were based on empirical data relating replicate ChIP-seq experiments (Extended Data Fig. 1b, c). Individual genome sequences for C57BL/6J and BALB/cJ were constructed in regions of interest using the reference (C57BL/6J) sequence and replacing BALB/cJ alleles at SNPs and indels reported in the vcf files from ref. 17.

Strain-specific motifs. De novo motif finding in ChIP-Seq-enriched regions from both mouse strains was used to define PWMs for transcription factors of interest (Extended Data Fig. 3a, b and Supplementary Table 1). These PWMs were used to define strain-specific motifs by using the options 'homer2 -find < individual genome sequence >' in HOMER for each genome sequence for the regions of interest. The positions of the identified motifs were compared between strains, taking into account shifts caused by indels relative to peak start coordinates and which DNA strand matched the identified motifs. Motifs with alignments only in one genome were considered strain specific.

PolyA-RNA-Seq. For each condition, RNA was isolated from 5×10^6 thioglycolate-elicited macrophages with Trizol LS, and 15 μg RNA was DNase-treated using TURBO DNase (Ambion) according to the manufacturer's instructions and ethanol-precipitated. PolyA-RNA was selected from 7 μg total RNA using the MicroPoly(A)Purist kit (Ambion), according to the manufacturer's instructions. Isolated RNA was hydrolysed in a total volume of 20 μl with 2 μl RNA fragmentation buffer (Ambion) for 10 min at 70 °C. The reaction was stopped with stop buffer, and buffer was exchanged to Tris, pH 8.5, using P30 size-exclusion columns (Bio-Rad). The fragmented RNA (30 ng) was 5′-decapped in a total volume of 21 μl containing 0.5 μl tobacco acid pyrophosphatase (TAP, Epicentre), 2 μl 10× TAP buffer and 1 μl SUPERase-IN, and incubated for 2 h at 37 °C. To dephosphorylate RNA 3′ ends, 0.5 μl 10× TAP buffer, 1.5 μl water, 0.5 μl of 0.25 M MgCl$_2$ (4.17 mM final; 1 mM EDTA for maximum phosphatase activity) and 0.5 μl of 10 mM ATP (0.2 μM final to protect PNK) where added, and the reaction was incubated with 1 μl PNK (Enzymatics) for 50 min at 37 °C. RNA fragments were 5′-phosphorylated by adding 10 μl 10× T4 DNA ligase buffer, 63 μl water and 2 μl PNK, and incubated for 60 min at 37 °C. RNA fragments were isolated using Trizol LS, precipitated in the presence of 300 mM sodium acetate and 2 μl glycoblue (Ambion), washed twice with 80% ethanol and dissolved in 4.5 μl water.

To prepare sequencing libraries, 0.5 μl of 9 μM 5′-adenylated 3′MPX adaptor/ 5Phos/AGATCGGAAGAGCACACGTCTGA/3AmMO/ (IDT, desalted; adenylated with Mth ligase (NEB) according to the manufacturer's instructions, phen–chloroform/chloroform-extracted, ethanol-precipitated with glycogen and dissolved in water at 9 μM) was heat-denatured together with the RNA for 2 min at 70 °C, and ligated with 100 U truncated T4 RNA ligase 2 K227Q (NEB) in 10 μl 1× T4 RNA ligase buffer without ATP, containing 10 U SUPERase-In and 15% PEG8000 for 2 h at 16 °C. To reduce adaptor dimer formation, 0.5 μl of 10 μM MPX_RT primer 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′ (IDT, desalted) was added and annealed to the ligation product by incubating at 75 °C for 2 min, then 37 °C for 30 min, and 25 °C for 15 min. Finally, 0.5 μl of 5 μM hybrid DNA/RNA sRNA 5′h adaptor 5′-GTTCAGAGTTCTACArGrUrCrCGr ArCrGrArUrC-3′ (IDT) was ligated to previously capped RNA 5′ ends by adding 2 μl T4 RNA ligase buffer, 6 μl 50% PEG8000 (15% final), 1 μl of 10 mM ATP, 9.5 μl water and 0.5 μl (5 U) T4 RNA ligase 1 for 90 min at 20 °C. To 15 μl of ligation reaction, an additional 0.5 μl of 10 μM MPX_RT primer was added, reactions were denatured at 70 °C for 1 min and then placed on ice. RNA was reverse-transcribed by adding 3 μl 10× first-strand buffer (4.5 μl water, 1.5 μl of 10 mM dNTP, 3 μl of 0.1 M DTT, 1.5 μl RNaseOUT and 1 μl Superscript III reverse transcriptase (Invitrogen)), and incubating for 30 min at 50 °C. Complementary DNA was isolated by adding 35 μl AMPure XL beads (Beckman), binding and washing according to manufacturer's instructions, and dissolving in 40 μl TET (0.1% Tween 20 in TE buffer). Libraries were PCR-amplified for 9 (polyA-RNA-Seq), 11 (5′-GRO-Seq), 12 (rRNA-5′-RNA-Seq) or 13 (polyA-5′-RNA-Seq) cycles, with 0.75 μM oNTI201 primer and TruSeq-compatible indexed primers (for example, 5′-CAAGCAGAAGANGACGGCATACGAGATnnnnnnNGTGACTGGA GTTCAGACGTGTGCTCTT-3′ (IDT, desalted; index in lowercase letters) using Phusion Hot Start II in Phusion HF buffer (Thermo Scientific) containing 0.5 M betaine (98 °C, 30 s; 12× (98 °C, 10 s; 57 °C, 25 s; 72 °C, 20 s); 72 °C, 1 min; stored at 4 °C, and 175–225-bp fragments were size-selected on 10% PAGE gels. Libraries were diluted 1:10^5 with TET buffer and quantified relative to samples of known cluster density by SYBR green qPCR with primers Solexa_1G_A 5′-AA TGATACGGCGACCACCGA-3′ and Solexa_1G_B 5′-CAAGCAGAAGACGG CATACGA-3′ (95 °C, 15 min; 25× (95 °C, 10 s; 60 °C, 60 s)), and sequenced for 51 (insert) +7 (index) cycles on a HiSeq 2000 sequencer (Illumina) with sRNA sequencing primer 5′-CGACAGGTTCAGAGTTCTACAGTCCGACGATC-3′ and TruSeq Index sequencing primer (Illumina).

GRO-Seq. GRO-Seq was performed as described previously[32] using 10^7 cells per condition. RNA at RefSeq transcripts was quantified for GRO-Seq and polyA-RNA-Seq by counting the normalized tags (to 10 million tags per experiment) in annotated exons for each RefSeq transcript.

Odds ratio calculations and statistical testing. Odds ratios for observing C57BL/6J-specific motif mutations relative to BALB/cJ-specific motif mutations in different classes of bound/modified loci (for example, Fig. 2b) were calculated using $(p_1/(1 - p_1))/(p_2/(1 \times p_2))$, in which p_1 is the frequency of C57BL/6J-specific motifs, and p_2 is the frequency of BALB/cJ-specific motifs. For Extended Data Fig. 4j, k, p_1 is the frequency of indicated events occurring in differentially bound loci and p_2 is the frequency in similarly bound loci. Unless otherwise indicated, t-tests were two-sided assuming unequal variance.

eQTL analysis. eQTL analysis was performed as previously described[23,25]. In brief, thioglycolate-elicited peritoneal macrophages were collected from 85 strains of mice. RNA was processed and hybridized to Affymetrix Genome HT_MG-430A. There were 22,416 probe sets analysed after removing individual probes overlapping SNPs and probe sets with 8 or more probes overlapping SNPs. Expression data was RMA normalized.

A total of 3,918,755 SNPs with a minimum minor allele frequency of 10% originating from mouse Perlegen variation data set[36] was imputed across the strains[37], and filtered to 3,695,041 SNPs based on proximity (<2 Mb) to transcription start sites of transcripts detectable by the microarray. Gene expression for each transcript was associated to SNPs within 2 Mb using the efficient mixed-model association mapping that corrects for population structure[38]. Association P values less than 1×10^{-5} (<1% FDR) were deemed significant[23]. The 3,695,041 SNPs used for association mapping were overlapped with H3K4me2 and H3K27ac regions. Because H3K4me2 and H3K27ac regions ranged from 500 to 1,500 bp whereas haplotype blocks averaged 300 kb, we considered SNPs outside H3K4me2/H3K27ac regions that were in linkage disequilibrium with a SNP in H3K4me2/H3K27ac regions as overlapping. Haplotype blocks were estimated in Haploview[39] using 143 strains with the following options: blockMAFThresh = 0.1, blockCutLowCI = 0.8, blockCutHighCI = 0.98, blockRecHighCI = 0.9, and blockInformFrac = 0.95. SNPs in linkage disequilibrium with enhancer SNP were considered markers of H3K4me2 regions. To test for enrichment of significant eQTL in H3K4me2 regions, we used the hypergeometric distribution function as follows:

$$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$$

in which k successes represents the number of significant eQTL in (or in linkage disequilibrium with) H3K4me2 regions; m denotes the number of SNPs with significant eQTL; N denotes the total number of SNPs; and n denotes the total number of SNPs in H3K4me2 regions.
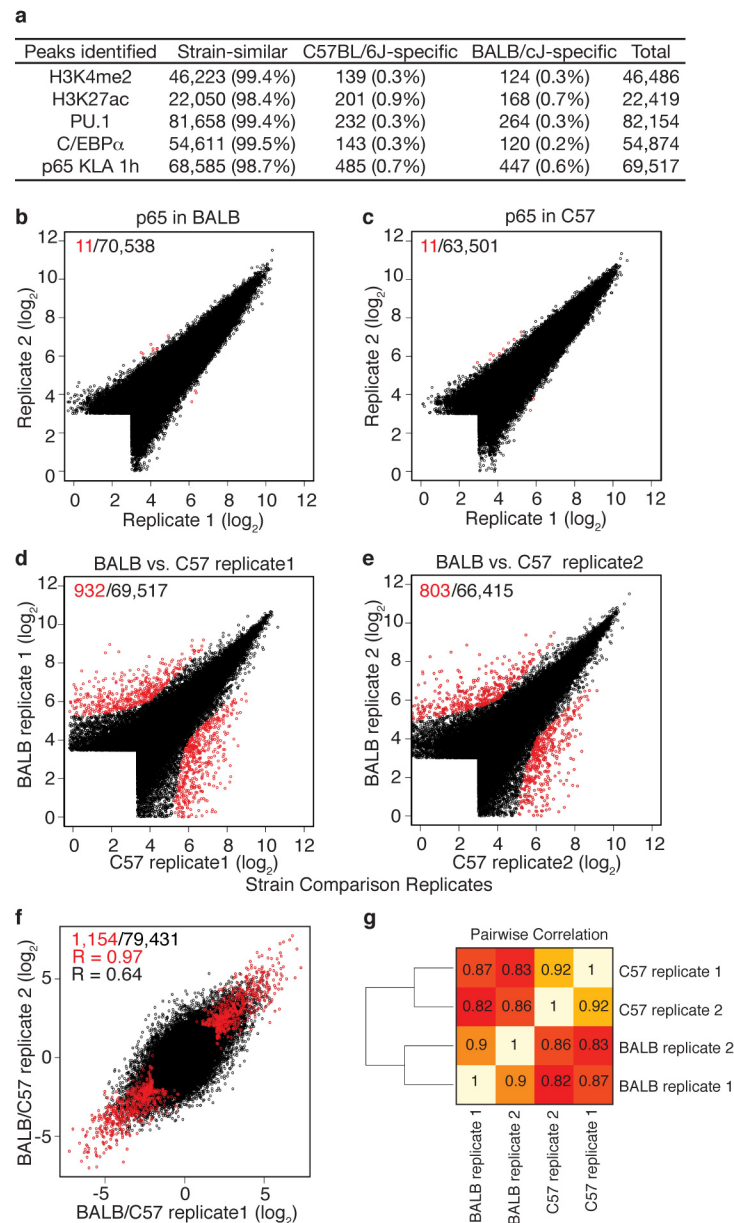
Macrophage eQTL enrichment in enhancers from other cell types. The short-read archive files were downloaded from the GEO under accession GSE24164 (ref. 20) for ChIP-sequencing for the H3K27ac mark in whole liver (Sequence Read Archive accession SRX027340), pro-B cells (SRX027345), and embryonic stem cells (SRX027331 and SRX027332), and input chromatin as background (liver: SRX027343, pro-B: SRX027348, stem cells: SRX027352). Sequencing reads were mapped to the C57BL/6J genome. H3K27ac regions were identified where tag pile-ups exceeded four times the input tags using HOMER[6], and interrogated for enrichment of significant macrophage eQTLs as described for macrophage H3K4me2 and H3K27ac regions above.

Reporter assays and mutation analysis. One-kilobase enhancers were PCR-amplified from C57BL/6J and BALB/cJ genomic DNA using genomic primers not overlapping variants that introduced terminal BamHI, BglII or BclI sites on one end and SalI or XhoI sites on the other end of the PCR products, depending on the restriction site content of the enhancer. These were digested with the respective restriction enzymes and ligated into a modified, BamHI- and SalI-digested pGL4.10 luciferase reporter plasmid (Invitrogen) containing a minimal HSV-TK promoter derived from pTAL-Luc (Clontech) (see Fig. 5a). Alternatively, 1-kb fragments were amplified using primers that introduced overhangs identical to the sequences flanking the BamHI/SalI tandem site in the pGL4.10 plasmid. Fragments were purified from the PCR reaction by SPRI using magnetic beads and cloned into the BamHI/SalI-cut reporter plasmid described above using Gibson Assembly master mix (NEB) according to the manufacturer's instructions. Mutations were introduced by PCR amplification with complementary primers containing the mutation to be introduced in the centre, followed by DpnI digestion of the template and transformation of bacteria. All constructs were confirmed by sequencing. For each reporter assay, 300 ng plasmid was transfected into RAW264.7 macrophages using SuperFect (Qiagen) together with 300 ng UB6 promoter-driven β-galactosidase reporter for transfection normalization in 24-well plates seeded with 1×10^5 cells 24 h before transfection. Twenty-four hours after transfection, media alone (RPMI plus 10% FBS) or media containing $100 \, \text{ng ml}^{-1}$ KLA was added for an additional 16 h. Luciferase activity was measured 24 h after transfection using a Veritas microplate luminometer (Turner Biosystems), and normalized to β-galactosidase activity (Applied Biosystem) for transfection efficiency. Each experiment was performed at least three independent times, with each reaction done in triplicates. Data represented as mean ± s.d., and statistical significance was determined by a one-sided t-test.

Stable transfected cell lines were made by transient co-transfection of the linearized reporter plasmids together with linearized neomycin resistance-expressing pcDNA3 vector as described above, followed by incubation with $275 \, \mu\text{g ml}^{-1}$ geneticin (G418 Sulphate, Invitrogen) for 2–3 weeks. Bulk cells from stably transfected colonies were tested for luciferase activity and normalized to total protein concentration (DC Protein Assay, BioRad).

33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
35. Hochberg, Y. B. Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. A Stat. Soc.* **57,** 289–300 (1995).
36. Frazer, K. A. et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448,** 1050–1053 (2007).
37. Kirby, A. et al. Fine mapping in 94 inbred mouse strains using a high-density haplotype resource. *Genetics* **185,** 1081–1095 (2010).
38. Kang, H. M. et al. Efficient control of population structure in model organism association mapping. *Genetics* **178,** 1709–1723 (2008).
39. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–265 (2005).

a

Peaks identified	Strain-similar	C57BL/6J-specific	BALB/cJ-specific	Total
H3K4me2	46,223 (99.4%)	139 (0.3%)	124 (0.3%)	46,486
H3K27ac	22,050 (98.4%)	201 (0.9%)	168 (0.7%)	22,419
PU.1	81,658 (99.4%)	232 (0.3%)	264 (0.3%)	82,154
C/EBPα	54,611 (99.5%)	143 (0.3%)	120 (0.2%)	54,874
p65 KLA 1h	68,585 (98.7%)	485 (0.7%)	447 (0.6%)	69,517

Extended Data Figure 1 | ChIP-Seq data characteristics. a, Summary of ChIP-Seq features identified. The number of ChIP-seq regions/peaks identified in untreated primary thioglycolate-elicited macrophages is tabulated for H3K4me2, H3K27ac, PU.1 and C/EBPα. Peaks for p65 were quantified in macrophages treated with $100\,\mathrm{ng\,ml}^{-1}$ KLA for 1 h. Unless otherwise noted, modification and binding were considered strain-specific at ≥fourfold difference between strains in sequenced tags, and the FDR was $<1\times10^{-14}$ based on Poisson cumulative distribution testing and Benjamini and Hochberg correction. b–e, Reproducibility and strain-specific binding. Two separate pools of thioglycolate-elicited macrophages from mice from each strain (C57BL/6J and BALB/cJ) were treated with KLA for 1 h. ChIP-seq for p65 was performed separately on each pool (see Methods). The number of normalized sequencing tags at the union of peaks identified in the indicated experiments is shown. Peaks highlighted in red were deemed experiment-specific using criteria applied throughout this study (fourfold, and FDR $<1\times10^{-14}$ from the cumulative Poisson distribution and Benjamini and Hochberg FDR estimation). The number of experiment-specific peaks is indicated (red) relative to the total number of peaks (black). f, Comparison of the p65 $\log_2$ peak tag ratio between strains and experimental sets for all peaks (black), highlighting experiment-specific peaks (red) identified in either d or e. g, Heat map showing pairwise correlation between all p65 experiments. Pearson correlation coefficients are given for each comparison.
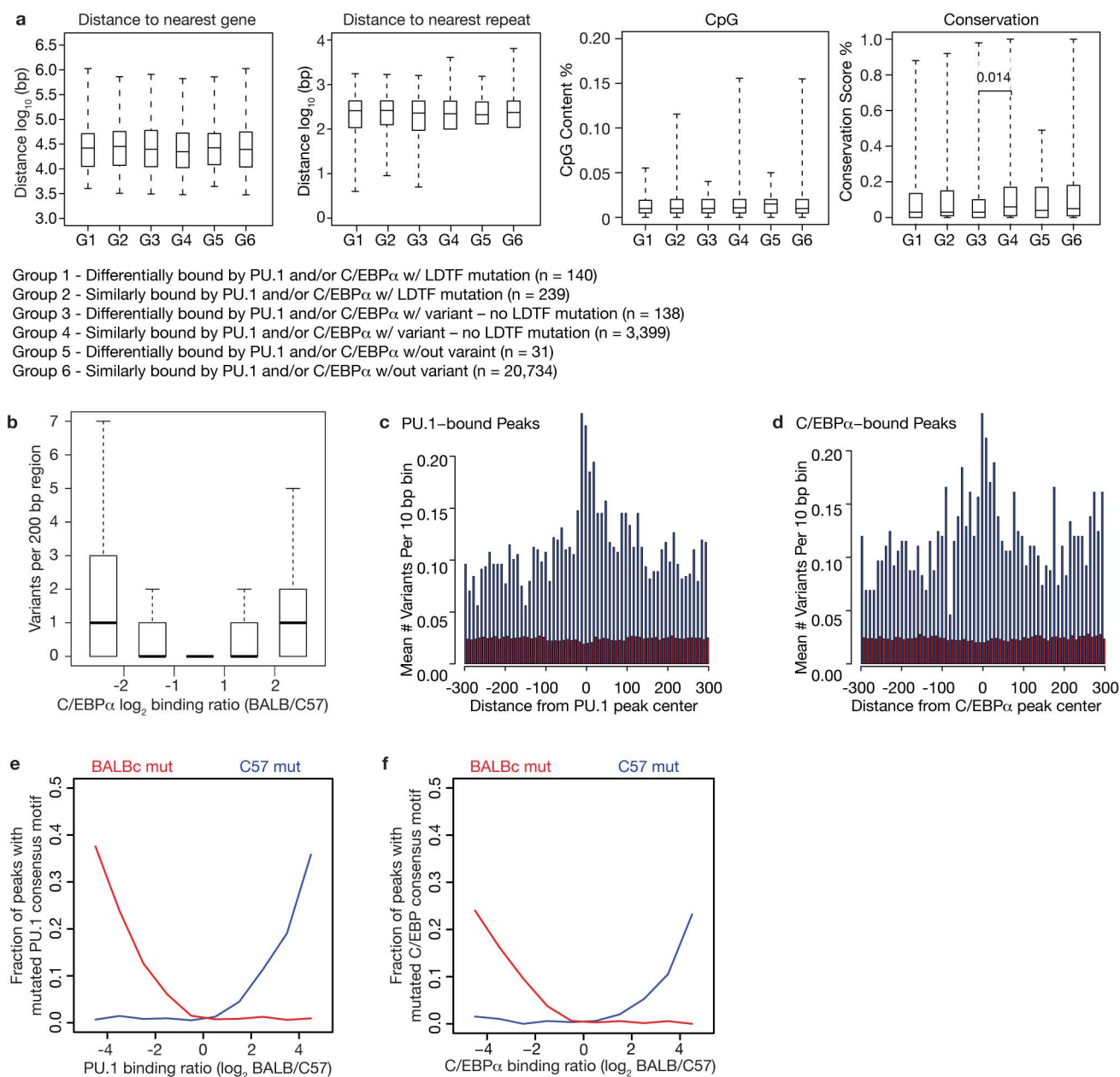
Extended Data Figure 2 | Strain-specific LDTF binding correlates with variant density and location in LDTF motifs but not with genomic context.
a, Genomic features do not distinguish between strain-similar and strain-specific LDTF binding. Peaks were restricted to promoter-distal peaks (>3 kb to gene start sites). Genomic features (distance to nearest gene, distance to nearest repeat, CpG content and conservation score) were compared among three pairs of strain-similarly bound and strain-specifically bound PU.1 and/or C/EBPα loci (listed as groups 1–6). Box midlines are medians, boundaries are first and third quartiles. Whiskers extend to the extreme data points. CpG content and conservation were quantified in 1-kb regions centred on the LDTF peak. P values from two-sided t-test are given if below 0.05. **b**, Strain-specific C/EBPα binding occurs in regions with increased variant density. ChIP-Seq tag

counts in 200-bp peak regions were stratified into five bins according to $\log_2$ ratios of peak tag counts in BALB/cJ versus C57BL/6J mice (x axis, $\log_2$ ratio), and the variant density distributions are shown per bin. **c, d**, Variant density distribution in strain-specific peaks. Mean variant densities within 10-bp bins relative to ChIP-Seq peak centres in strain-similar (red) or strain-specific (blue) peaks. **e**, Strain-specific PU.1 binding correlates with mutations in PU.1 motifs. PU.1 motif mutations were quantified in PU.1-bound regions and plotted against the logarithmic ratio of PU.1 peak tag counts in each strain (binding ratio) (x axis). The frequency of motifs that were mutated in BALB/cJ are plotted in red and those mutated in C57BL/6J in blue. **f**, The analogous relationship as shown in **e** for PU.1 is plotted for C/EBP motif mutations versus C/EBPα strain binding ratio.
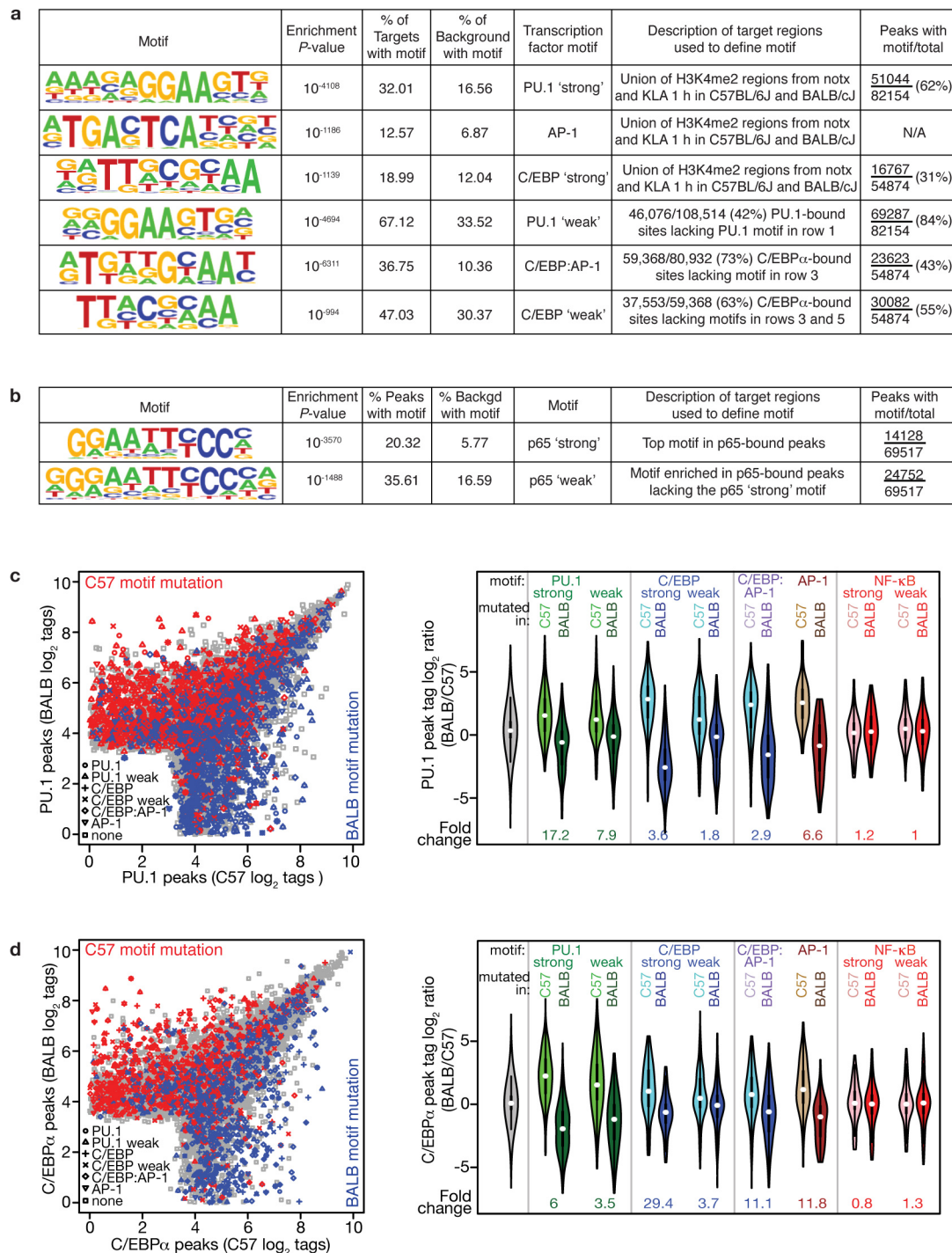
Extended Data Figure 3 | Strain-specific PU.1 and C/EBPα binding correlates with strain-specific LDTF motifs. a, Top and degenerate motifs enriched in H3K4me2 and PU.1 or C/EBPα ChIP-Seq peaks. b, NF-κB consensus and degenerate motifs enriched in p65 ChIP-Seq peaks. These motifs were used to query individual genome sequences and identify strain-specific motifs in subsequent analysis. Degenerate 'weak' motif occurrence numbers for a given factor include ChIP-Seq peaks containing 'strong' motifs. Position weight matrices and log-odds score thresholds for each motif are given in Supplementary Table 1. c, d, Mutations in LDTF motifs affect PU.1 (c) and C/EBPα (d) binding. Left panels show scatterplots for the ChIP-Seq-defined binding of PU.1 (c) and C/EBPα (d) between C57BL/6J (x axes) and BALB/cJ (y axes). Strain-specific motifs were queried within 100-bp of each peak position. Red symbols designate binding events at loci where a

polymorphism mutated a C/EBP, PU.1 or AP-1 motif in the C57BL/6J genome, whereas the motif was intact in the BALB/cJ genome. Blue points highlight mutations in these motifs in the BALB/cJ genome only. Violin plots in the right panels show the effect of each motif mutation (along x axes: PU.1, C/EBP, AP-1 and NF-κB) on the ratio of PU.1 (c) and C/EBPα (d) binding between mouse strains, (y axes: positive values denote BALB/cJ-specific, negative values denote C57BL/6J-specific). Tag ratio distributions for peaks overlapping C57BL/6J motif mutations are on the left (light colours), those for peaks overlapping BALB/cJ motif mutations are on the right (dark colours). The fold-difference between mean binding ratios is indicated under the pair of distributions for each motif. The grey distribution indicates PU.1- or C/EBPα-bound loci not overlapping strain-specific motifs.
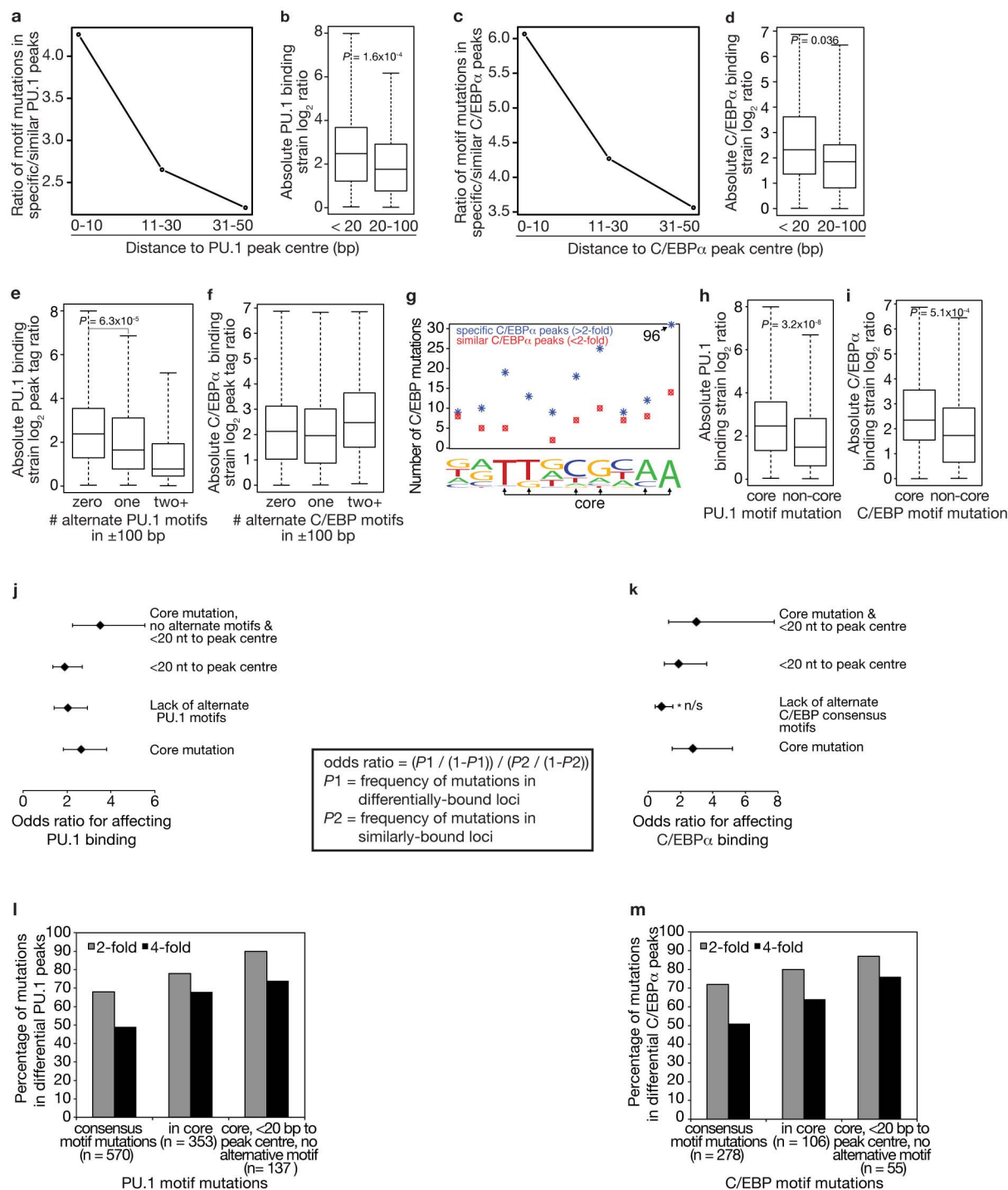
Extended Data Figure 4 | Effects of cognate motif distance from peak centre, variant position within a motif and the presence of alternative motifs on strain-differential binding of PU.1 and C/EBPα. **a–d,** PU.1 and C/EBP motif mutations near the experimentally derived peak centre are associated with impaired binding. **a, c,** The ratios of the frequencies of variant-containing motifs at the given distances from strain-differentially versus strain-similarly bound peak centres (>twofold versus <twofold tag count ratio) for 570 PU.1 (**a**) and 278 C/EBP (**b**) variant-containing motifs are shown, respectively. **b, d,** The distribution of absolute strain peak tag count ratios of peaks whose centre is at the given distances from mutated PU.1 (**b**) or C/EBP (**d**) motifs. Box midlines are medians, and boundaries are first and third quartiles. Whiskers extend to the extreme data points. *P* values are from two-sided *t*-test. **e, f,** Effects of alternative PU.1 and C/EBP motifs and core mutations on binding. The number of non-mutated 'alternative' PU.1 or C/EBP motifs in the strain with a PU.1 or C/EBP motif mutation was counted, and the absolute respective PU.1

or C/EBPα log$_2$ strain binding ratio is shown. **g,** Defining the C/EBP motif core by comparing differential versus similar C/EBPα binding. Sequence variants within C/EBP motifs located in loci devoid of alternative C/EBP motifs (*n* = 178) were counted according to whether they were in differential (blue) or similar (red) C/EBPα-bound peaks. **h,** The distribution of PU.1 binding strain log$_2$ ratios (*x* axis) is shown for PU.1 mutations located in the PU.1 core and non-core nucleotides (defined in Fig. 1g). **i,** The C/EBPα binding strain log$_2$ ratio is shown for C/EBP core and non-core mutations as defined in **g**. **j, k,** Motif mutations predominately occur at differentially bound loci. The odds ratios (*x* axis; equation shown in box) describing the relative effect of the indicated characteristics of mutated motifs on differential binding relative to similar binding are shown for PU.1 (**j**) and C/EBPα (**k**). Whiskers show 95% confidence intervals. nt, nucleotides. **l, m,** The percentage of respective motif mutations consistent with altered PU.1 (**l**) and C/EBPα (**m**) binding is shown for the indicated categories of motif mutations.
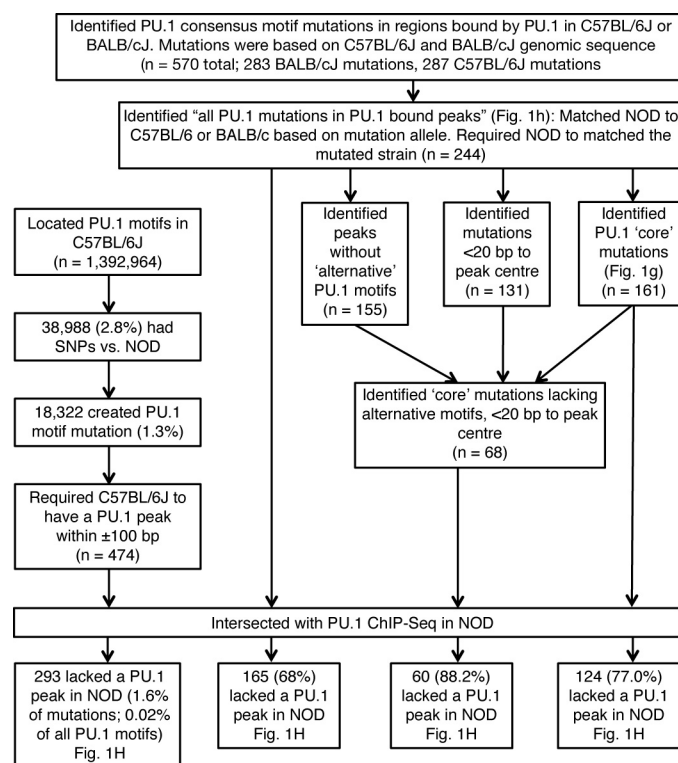
Extended Data Figure 5 | Analysis pipeline for predicting functional PU.1 mutations in NOD. Data are shown in Fig. 1H.
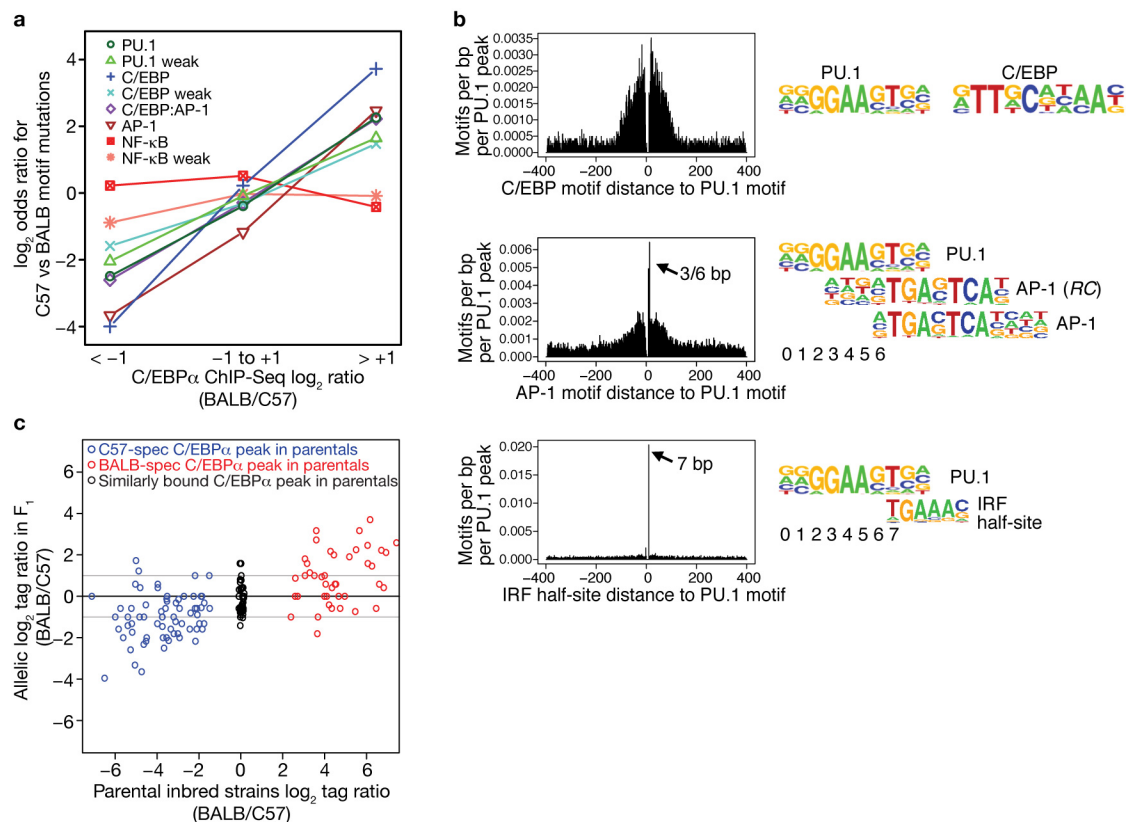
Extended Data Figure 6 | LDTF motif mutations are enriched at strain-specific C/EBPα-bound loci relative to strain-similar loci. a, The $\log_2$ odds ratio for observing a C57BL/6J-specific versus BALB/cJ-specific mutation in the indicated three bins of C/EBPα binding ratios: similar (middle bin), or strain-specifically C/EBPα bound (left and right bins). Details are in the Methods. **b,** Collaborative binding is largely not mediated by direct protein–protein interactions. A total of 14,199 loci bound by PU.1 and C/EBPα were centred on the PU.1 weak motif (0 on x axes) and cumulative instances of C/EBP and AP-1 motifs were plotted at each position relative to the central PU.1 motif. Interferon response factor (IRF) half-sites are plotted as control for a factor that requires direct protein–protein interactions with PU.1 for DNA binding. The motifs in each comparison showing overlapping sequence and base pair

distances are indicated to the right. Peak distances from the central PU.1 motif are indicated in the histograms. RC denotes reverse complement. **c,** Allele-specific C/EBPα binding in F_1 heterozygotes is similar to binding in homozygous parental strains. C/EBPα ChIP-seq reads from CB6F1/J hybrid F_1 macrophages were mapped with no mismatches to both parental genome sequences to identify allele-specific reads. C/EBPα $\log_2$ peak tag ratios between the parental strains (BALB/cJ versus C57BL/6J) are shown on the x axis, and the $\log_2$ ratio of allele-specific reads in the F_1 hybrids are shown on the y axis (BALB/cJ allele versus C57BL/6J allele). C57BL/6J-specific C/EBPα regions are blue, BALB/cJ-specific C/EBPα regions are red, and strain-similar C/EBPα regions are black. Strain-specific or similar regions were defined from parental data.
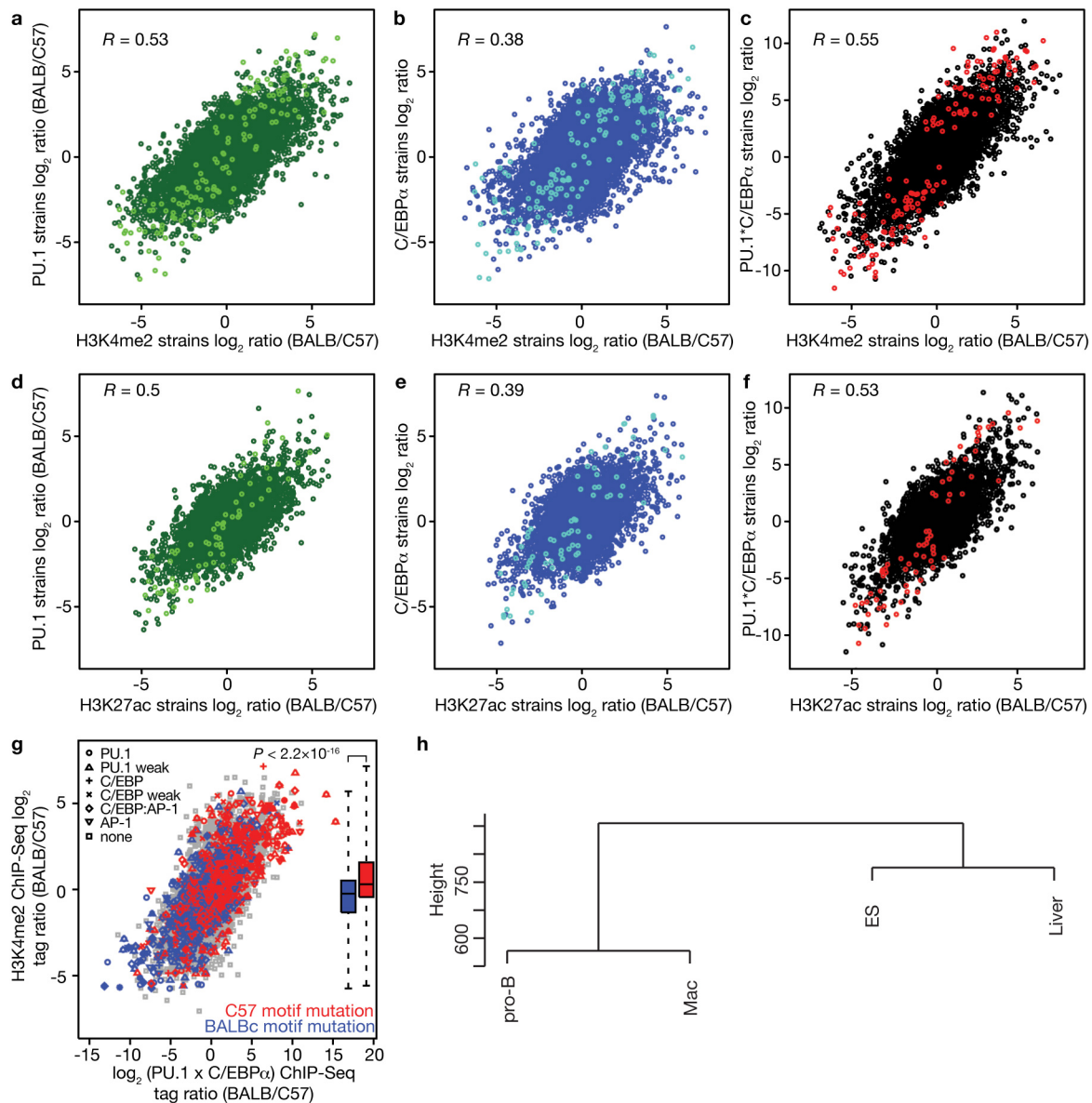
Extended Data Figure 7 | Strain-specific epigenetic marks correlate with LDTF binding, and LDTF mutations segregate with altered H3K4me2 deposition. a–f, Strain-specificity of LDTF binding and epigenetic marks. The relative amount of H3K4me2 (a–c) and H3K27ac (d–f) between C57BL/6J and BALB/cJ (x axes) is highly correlated with the amount of bound PU.1, C/EBPα or product (PU.1 × C/EBPα). The $\log_2$ ratios of the peak tag counts for PU.1, C/EBPα and PU.1 × C/EBPα in each strain are shown relative to the $\log_2$ of the peak tag count ratios for H3K4me2 or H3K27ac. Loci containing strain-specific LDTF motifs in a differentially PU.1- or C/EBPα-bound peak are highlighted. Correlation coefficients (Pearson) are indicated for each comparison. g, LDTF mutations segregate with altered H3K4me2 deposition. The $\log_2$ of the ratio of the product of the normalized peak tag counts for PU.1 and C/EBPα in 200 bp in each strain (x axis) is compared to the $\log_2$ H3K4me2 peak tag ratio in 1 kb (y axis) for loci containing at least a PU.1 or C/EBPα peak. Strain-specific LDTF motif mutations are indicated by the designated symbols and coloured by the mutated strain (C57BL/6J red, BALB/cJ blue). The distribution of H3K4me2 strain ratios stratified by corresponding LDTF strain mutations is shown to the right, with P value from a two-sided t-test. h, Relationships between H3K27ac patterns in different cell types. ES, embryonic stem. Hierarchical clustering of H3K27ac-positive regions as determined by ChIP-Seq and analysis with HOMER. The number of ChIP-seq tags in each of the 86,264 H3K27ac-marked regions used for comparison with eQTL data in Fig. 2e that were detected in at least one cell type was clustered using Euclidean distance.
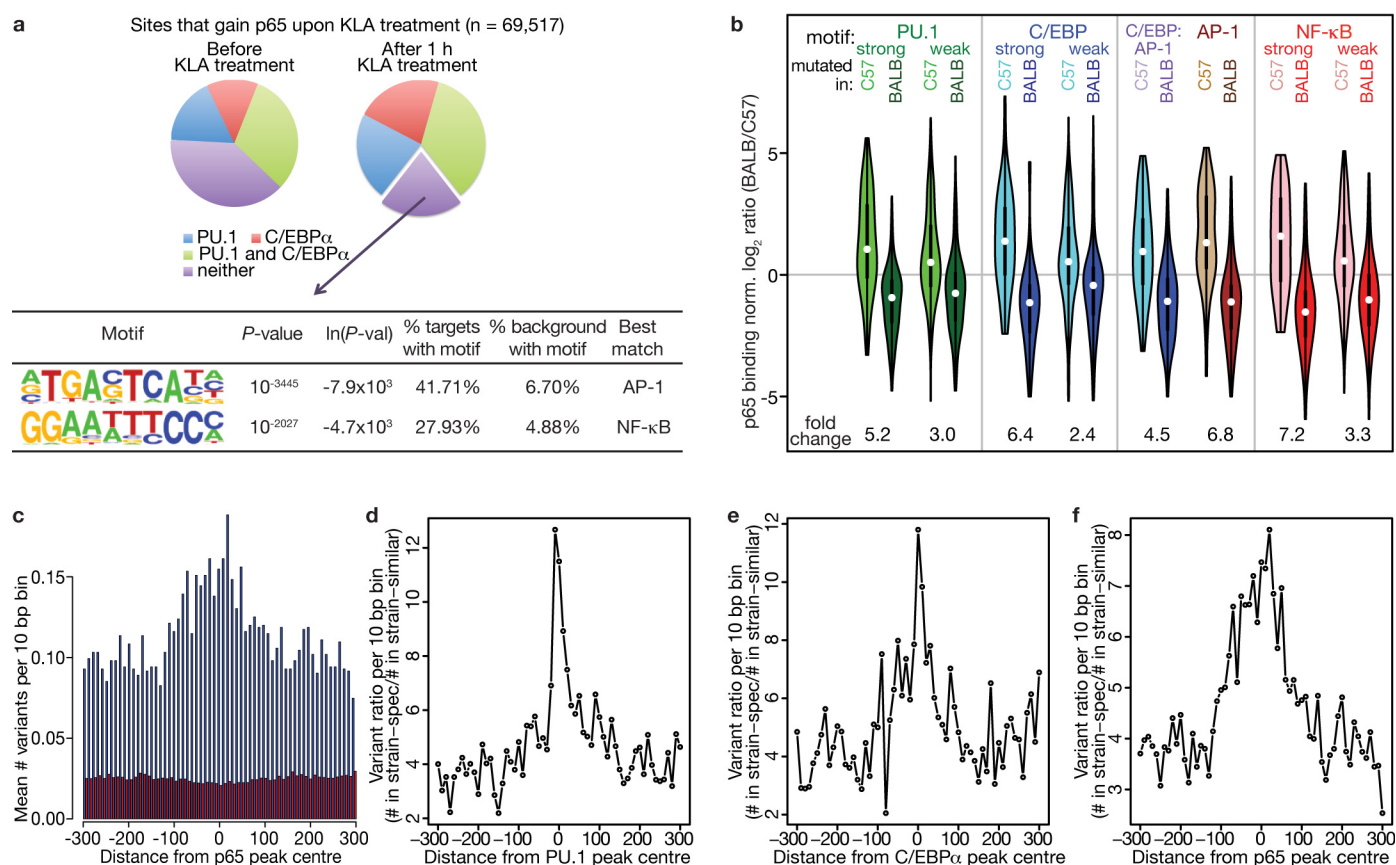
Extended Data Figure 8 | LDTFs prime the p65 cistrome. a, The 69,517 regions that gained p65 in C57BL/6J after KLA treatment were analysed for binding of PU.1 and C/EBPα with and without KLA treatment as shown in the pie charts. Loci not bound by PU.1 or C/EBPα after KLA treatment were analysed by *de novo* motif finding. The most enriched motif was AP-1, and the second-most enriched motif was NF-κB. **b**, Violin plots of the p65 strain ratios of mean-normalized p65 binding for p65-bound peaks stratified by motifs mutations present in either BALB/cJ or C57BL/6J. Mutated motifs included PU.1 (strong and weak), C/EBP (strong and weak), C/EBP:AP-1 heterodimers, AP-1 and NF-κB. The effect on p65 binding per group is shown by comparing the mean-normalized p65 tag binding ratio along the *y* axis (log$_2$(BALB/cJ–C57BL/6J); positive values denote BALB/cJ-specific, negative

values denote C57BL/6J-specific). White circles indicate the distribution means, and the average fold change associated with C57BL/6J-mutating and BALB/cJ-mutating SNPs in the respective motifs is given beneath. One-sided *t*-test *P* values between each pair of distributions ranged from 1×10^{-29} to 1×10^{-14}. **c**, Variant density in strain-specific and strain-similar p65 peaks. Mean variant density within 10-bp bins relative to p65 ChIP-Seq peak centres in strain-similar (red) or strain-specific peaks (blue). **d–e**, The variant density distribution in strain-specific p65 peaks is broader than those for PU.1 or C/EBPα. Fold enrichment of variant densities in strain-specific relative to strain-similar peaks (*y* axes) for PU.1 (**d**), C/EBPα (**e**) and p65 (**f**) is shown relative to the peak centres (*x* axes). Ratios plotted in **d** and **e** are from data in Extended Data Fig. 2c and d, respectively.
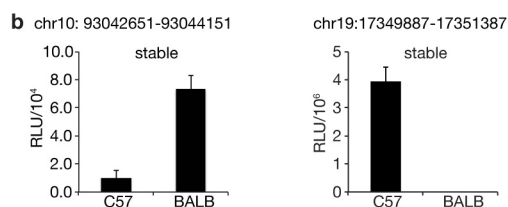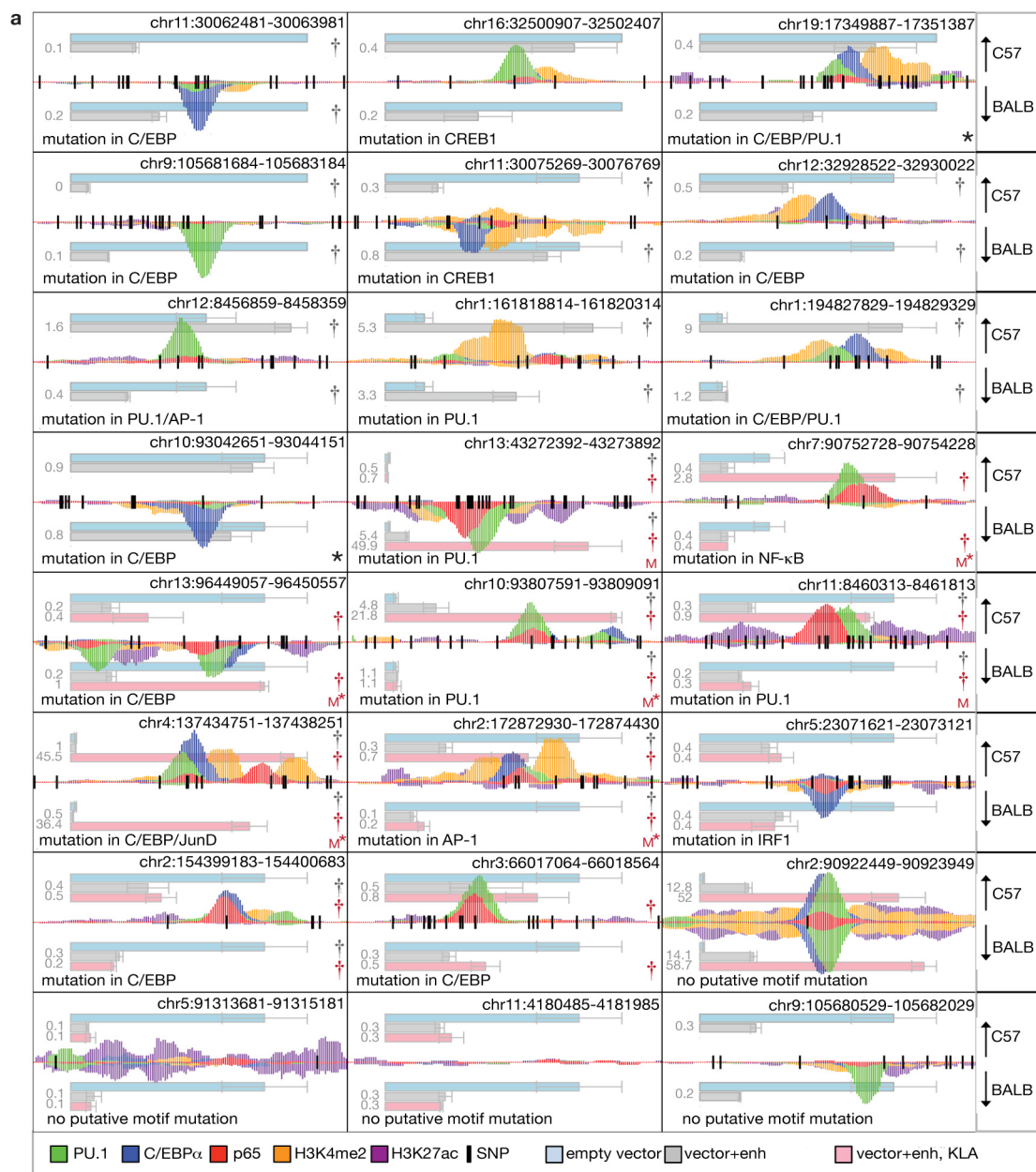
Extended Data Figure 9 | Validation of strain-specific enhancer activity.
a, Enhancer activity in transient reporter assays correlates with strain-specific
LDTF and p65 binding. Luciferase assay results for 24 loci (20 strain-specific
enhancers with strain-specific motifs, 1 positive control with strain-similar
enhancer activity (row 7, column 3), 2 negative controls lacking enhancer
activity in both strains (row 8, columns 1 and 2), and 1 strain-specific enhancer
lacking a strain-specific motif (row 8, column 3)) in transiently transfected
RAW264.7 cells 48 h after transfection. Each 1-kb locus is represented by the
horizontal midline within a box (see Fig. 5). ChIP-seq tag pile-ups are shown for
PU.1 (green), C/EBPα (blue), p65 (red), H3K27ac (purple) and H3K4me2
(orange) for C57BL/6J (above midline) and BALB/cJ (below midline) with
identical scales. Binding/modification data are shown after treatment with
100 ng ml^{-1} KLA. Vertical black lines indicate SNP locations. Horizontal bars
indicate average luciferase (enhancer) activity of the empty vector (blue, no
enhancer), activity of a locus cloned from either strain in grey C57BL/6J (above)
and BALB/cJ (below) under basal conditions, or after overnight stimulation
with 100 ng ml^{-1} KLA (pink). Luciferase values from transiently transfected
cells were normalized to the activity measured for a co-transfected UB6
promoter-β-galactosidase reporter construct. Empty vector values were scaled
to 0.5 for the first four loci, and to 1 for the remaining loci. Constructs in which
the predicted motif-disrupting variant alleles were swapped are denoted by 'M',
with mutations causing a significant effect in at least two out of three replicates
being denoted by an additional asterisk ($P < 0.05$, one-sided t-test). Error bars
show s.d. from three biological replicates, average values are indicated next to
each bar. Experiments were replicated at least three times. Significant strain-
specific enhancer activity is indicated by a dagger (grey without treatment, red
after KLA treatment, one-tailed t-test, $P < 0.05$). **b**, Chromatinization is
necessary for the strain specificity of a subset of enhancers. RAW264.7 cells
were stably transfected with the two constructs containing the loci that showed
strain-specific binding but lacked strain-specific enhancer activity in transient
reporter assays (row 4, column 1 and row 1, column 3, marked by an asterisk).
Luciferase activity measured in lysates of stably transfected cells was
normalized to total protein content. RLU, relative light units.

a Relative enhancer activity chart with legend: Fold C57 wt/BALB wt, Fold C57 mut to BALB/C57 wt, Fold BALB mut to C57/BALB wt.

Single SNP in LDTF; Multiple SNPs; LDTF variant swapped

C = motif mutation in C57BL6
B = motif mutation in BALB

b chr13: 96449307–96450307 — C57 putative C/EBP motif mutation; BALB

Legend: PU.1, C/EBPα, p65, SNP, H3K4me2, H3K27ac

chr7: 90752985–90753972 — C57; BALB putative NF-κB motif mutation

C/EBP motif — C57 TACTTGTCAA / BALB TATTTCTCAA
NF-κB motif — C57 GGAATTCCCT / BALB TGAATTCCCT

Wild type / Swap C/EBP SNP — Treatment: None, KLA
Wild type / Swap NF-κB SNP

c chr10:93807591–93809091 — C57 putative PU.1 motif mutation; BALB

Wild type / Swap 5' SNP / Swap PU.1 SNP / Swap 3' SNP — None, KLA

C57BL6 seq: TACTTTTTTAGT 154bp CAACAGAAGAGGAAATGGAAAA 82bp AAATGGTAAACA
BALBc seq: TACTTGTTTAGT CAACAGAAGAGGGAATGGAAAA AAATGATAAACA

5' SNP | PU.1 SNP | 3' SNP

Extended Data Figure 10 | Motif analysis identifies causal SNPs in enhancers. Regions of ~1 kb size centred on PU.1 or C/EBPα ChIP-Seq peaks of similar tag count in C57BL/6J and BALB/cJ (<twofold difference) that contain a variant in a motif for the respective factor within 100 bp of the peak centre were cloned into a luciferase reporter plasmid containing a minimal HSV-TK derived promoter. Three independent transient transfection experiments were performed in RAW264.7 cells, with triplicate transfections of each construct. Where indicated, variant nucleotides in a motif were mutated to that present in the other strain, and the resulting enhancer activity was scored relative to the wild-type allele. Shown are the ratios of the normalized luciferase activity of the C57BL/6J versus BALB/cJ alleles from a representative experiment. Luciferase values from transiently transfected cells were normalized to the activity measured for a co-transfected UB6 promoter β-galactosidase reporter construct. Error bars represent derived s.d. calculated by Gaussian error propagation. Constructs exhibiting significantly different activity ratios in two out of three experiments as determined by two-sided *t*-test ($P < 0.05$) are marked with an asterisk. Strain and motif mutated by a variant are indicated below denoted by the 'm' prefix. In the table, plus signs indicate whether a tested enhancer contains an alternative motif for the same factor, a variant at a motif position that is not located at a motif core as defined in Fig. 1g and Extended Data Fig. 4g, or a variant in a motif located less than 20 bp away from the peak centre. Characteristics of the loci and primer sequences are in Supplementary Table 3. **b**, Identifying causal variants by motif analysis. Left panels show the ChIP-Seq pile-ups and SNP locations as in Extended Data Fig. 9. Right panels plot the relative enhancer reporter luciferase activities of the loci shown on the left, either in the wild-type configuration or when swapping the SNP indicated by a black triangle by site-directed mutagenesis. Motifs mutated by the indicated SNPs are shown above, with the mutation underlined and in red. **c**, To confirm that the centrally located PU.1 motif is essential for the C57BL/6J-specific activity, a 1-kb fragment of the locus from C57BL/6J or BALB/cJ was cloned into the luciferase reporter as described in Fig. 5 and the effects of swapping alleles at the predicted causal PU.1 SNP and flanking control 5′ and 3′ SNPs on enhancer activity are shown. Swapping alleles at the PU.1 SNP reversed strain-specific enhancer activity, whereas swapping alleles at either flanking SNP had no significant effect.

ARTICLE

A small–molecule AdipoR agonist for type 2 diabetes and short life in obesity

Miki Okada-Iwabu[1,2,3]*, Toshimasa Yamauchi[1,2,3]*, Masato Iwabu[1,2]*, Teruki Honma[4], Ken-ichi Hamagami[1], Koichi Matsuda[1], Mamiko Yamaguchi[1], Hiroaki Tanabe[4], Tomomi Kimura-Someya[4], Mikako Shirouzu[4], Hitomi Ogata[5], Kumpei Tokuyama[5], Kohjiro Ueki[1], Tetsuo Nagano[6], Akiko Tanaka[4,6], Shigeyuki Yokoyama[4,7] & Takashi Kadowaki[1,2,3]

Adiponectin secreted from adipocytes binds to adiponectin receptors AdipoR1 and AdipoR2, and exerts antidiabetic effects via activation of AMPK and PPAR-α pathways, respectively. Levels of adiponectin in plasma are reduced in obesity, which causes insulin resistance and type 2 diabetes. Thus, orally active small molecules that bind to and activate AdipoR1 and AdipoR2 could ameliorate obesity-related diseases such as type 2 diabetes. Here we report the identification of orally active synthetic small-molecule AdipoR agonists. One of these compounds, AdipoR agonist (AdipoRon), bound to both AdipoR1 and AdipoR2 *in vitro*. AdipoRon showed very similar effects to adiponectin in muscle and liver, such as activation of AMPK and PPAR-α pathways, and ameliorated insulin resistance and glucose intolerance in mice fed a high-fat diet, which was completely obliterated in AdipoR1 and AdipoR2 double-knockout mice. Moreover, AdipoRon ameliorated diabetes of genetically obese rodent model *db/db* mice, and prolonged the shortened lifespan of *db/db* mice on a high-fat diet. Thus, orally active AdipoR agonists such as AdipoRon are a promising therapeutic approach for the treatment of obesity-related diseases such as type 2 diabetes.

The number of overweight individuals worldwide has grown markedly, leading to an escalation of obesity-related health problems associated with increased morbidity and mortality. Insulin resistance is a common feature of obesity and predisposes the affected individuals to a variety of pathologies, including type 2 diabetes and cardiovascular diseases. Although considerable progress has been made in understanding the molecular mechanisms underlying insulin resistance and type 2 diabetes, their satisfactory treatment modalities remain limited[1–4].

Adiponectin (*Adipoq*)[5–8] is an antidiabetic and antiatherogenic adipokine. Plasma adiponectin levels are decreased in obesity, insulin resistance and type 2 diabetes[9]. Replenishment of adiponectin has been shown to ameliorate insulin resistance and glucose intolerance in mice[10–12]. This insulin sensitizing effect of adiponectin seems to be mediated, at least in part, by an increase in fatty-acid oxidation via activation of AMP-activated protein kinase (AMPK)[13–15] and also via peroxisome proliferator-activated receptor (PPAR)-α[16,17].

We previously reported the expression cloning of complementary DNA encoding adiponectin receptors 1 and 2 (*Adipor1* and *Adipor2*)[18]. AdipoR1 and AdipoR2 are predicted to contain seven-transmembrane domains[18], but to be structurally and functionally distinct from G-protein-coupled receptors[19]. AdipoR1 and AdipoR2 serve as the major receptors for adiponectin *in vivo*, with AdipoR1 activating the AMPK pathways and AdipoR2 activating the PPAR-α pathways[20].

In skeletal muscle[21], AdipoR1 is predominantly expressed and activates AMPK[22] and PPAR-γ coactivator (PGC)-1α (ref. 23) as well as Ca^{2+} signalling pathways, which have also been shown to be activated by exercise[24,25]. Exercise has been reported to have beneficial effects on obesity-related diseases such as type 2 diabetes, and could contribute to healthy longevity[26]. Liver expresses AdipoR1 and AdipoR2, both of which have roles in the regulation of glucose and lipid metabolism,

inflammation, and oxidative stress *in vivo*[20]. Here we report the discovery of an orally active synthetic small molecule that binds to and activates both AdipoR1 and AdipoR2, ameliorates insulin resistance and type 2 diabetes, and prolongs the shortened lifespan of *db/db* mice.

Identification of small-molecule agonists of AdipoR

To identify orally active compounds that could bind to and activate AdipoR, we screened a number of small molecules in the chemical library at Open Innovation Center for Drug Discovery, The University of Tokyo[27]. We performed functional assays to determine the ability of small molecules to activate AMPK (Extended Data Table 1 and Extended Data Fig. 1) and to ascertain the dependency of small molecules on AdipoR in C2C12 myotubes by testing the effects of suppression of AdipoR expression by specific short interfering RNA (siRNA) on phosphorylation of AMPK stimulated with each compound (Extended Data Table 2 and Extended Data Fig. 2). We named one of these hits AdipoR agonist (AdipoRon; Fig. 1a). We also used compounds 112254 and 165073 in some of the experiments as another hit and a non-hit, respectively (Extended Data Tables 1 and 2 and Extended Data Figs 1 and 2).

The treatment of C2C12 myotubes with AdipoRon caused an increase in the phosphorylation of Thr 172 in the α-subunit of AMPK (αAMPK)[28]. AdipoRon at concentrations of 5–50 μM increased AMPK phosphorylation in a dose-dependent manner to almost the same extent as did adiponectin (Fig. 1b, c) without mitochondrial complex I inhibition (Extended Data Fig. 3a). Suppression of AdipoR1 by specific siRNA (Extended Data Fig. 3b, c) greatly reduced the increase in AMPK phosphorylation induced by AdipoRon (Fig. 1c), indicating that AdipoRon increased AMPK phosphorylation via AdipoR1. Compound number 112254 (another hit) also significantly increased phosphorylation of

[1]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan. [2]Department of Integrated Molecular Science on Metabolic Diseases, 22nd Century Medical and Research Center, The University of Tokyo, Tokyo 113-0033, Japan. [3]Department of Molecular Medicinal Sciences on Metabolic Regulation, 22nd Century Medical and Research Center, The University of Tokyo, Tokyo 113-0033, Japan. [4]RIKEN Systems and Structural Biology Center, Tsurumi, Yokohama 230-0045, Japan. [5]Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba 305-8577, Japan. [6]Open Innovation Center for Drug Discovery, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [7]Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan.
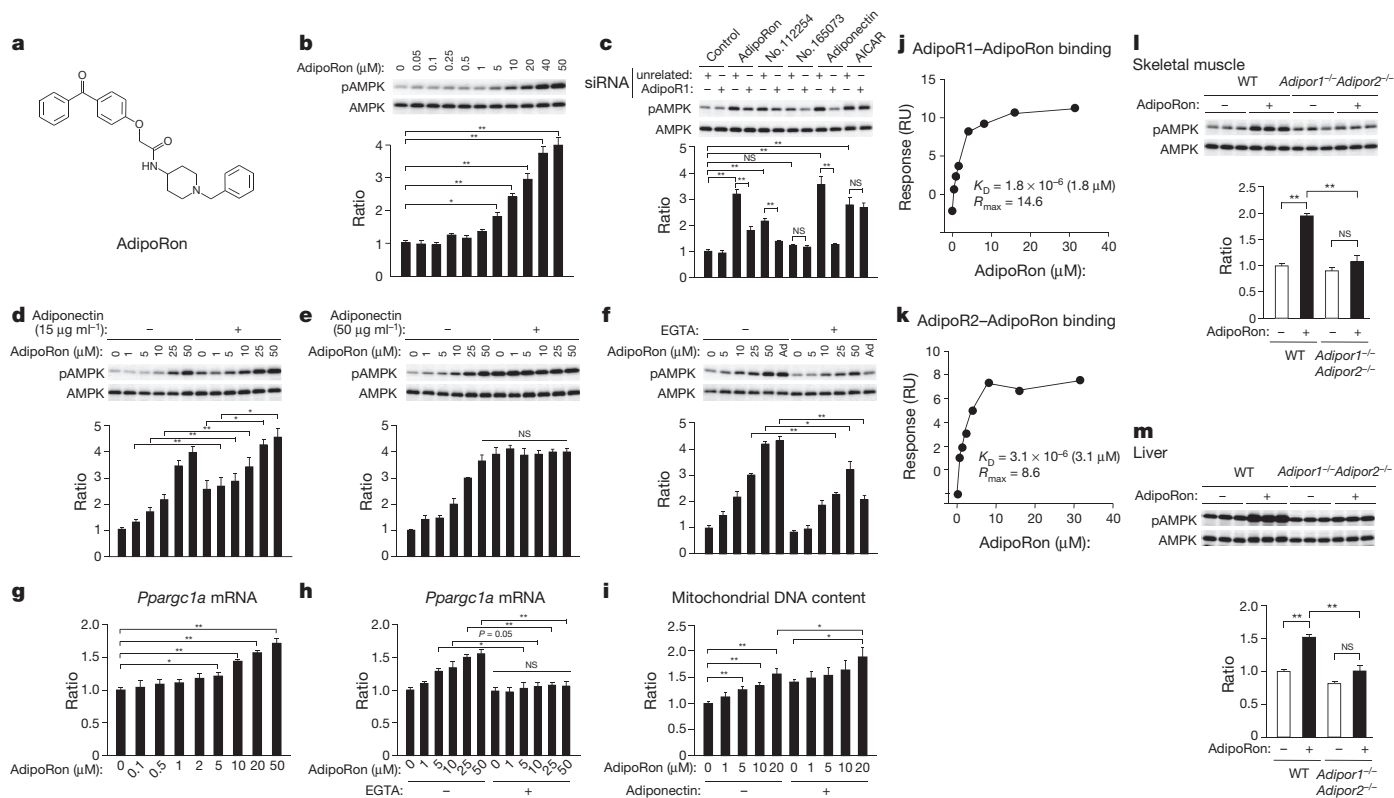*These authors contributed equally to this work.

Figure 1 | Small-molecule AdipoR agonist AdipoRon binds to both AdipoR1 and AdipoR2, and increases AMPK activation, PGC-1α expression and mitochondrial biogenesis in C2C12 myotubes. a, Chemical structure of AdipoRon. **b–i, l, m**, Phosphorylation and amount of AMPK (**b–f, l, m**), *Ppargc1a* mRNA levels (**g, h**), and mitochondrial content as assessed by mitochondrial DNA copy number (**i**) in C2C12 myotubes after myogenic differentiation (**b–i**), in skeletal muscle (**l**) or in liver (**m**) from wild-type (WT) or *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout mice, treated with indicated concentrations of AdipoRon (**b, d–i**) or adiponectin (**d**, 15 μg ml^{-1}; **e**, 50 μg ml^{-1}; **i**, 10 μg ml^{-1}), for 5 min (**b, d–f**), 1.5 h (**g, h**) and 48 h (**i**), with or without EGTA (**f, h**), 25 μM AdipoRon, compound 112254 and 165073, 30 μg ml^{-1} adiponectin for 5 min or 1 mM AICAR for 1 h and transfected with or without the indicated siRNA duplex (**c**), or AdipoRon (**l, m**). **j, k**, Surface plasmon resonance measuring AdipoRon binding to AdipoR1 and AdipoR2. AdipoR1 and AdipoR2 were immobilized onto a sensor chip SA. Binding analyses were performed using a range of AdipoRon concentrations (0.49–31.25 μM). All values are presented as mean ± s.e.m. **b, c, e–l**, $n = 4$ each; **d, l, m**, $n = 3$ each; *$P < 0.05$ and **$P < 0.01$ compared to control or unrelated siRNA or as indicated. NS, not significant.

AMPK via AdipoR1, albeit less potently, and compound 165073 (a non-hit) failed to increase phosphorylation of AMPK (Fig. 1c).

In the presence or absence of the submaximal concentration of adiponectin (15 μg ml^{-1}), AdipoRon increased AMPK phosphorylation in a dose-dependent manner (Fig. 1d), whereas AdipoRon did not increase nor decrease AMPK phosphorylation in the presence of the maximal concentration of adiponectin (50 μg ml^{-1}) (Fig. 1e). These data suggested that AdipoRon replenished AMPK phosphorylation stimulated by adiponectin.

EGTA partially suppressed the AdipoRon-induced increase in AMPK phosphorylation in C2C12 myotubes (Fig. 1f), indicating that extracellular free Ca^{2+} is required for full AMPK phosphorylation stimulated with AdipoRon, like adiponectin[21]. Moreover, AdipoRon increased PGC-1α (*Ppargc1a*) expression (Fig. 1g, h) and mitochondrial DNA content (Fig. 1i) in a dose-dependent manner. Furthermore, EGTA effectively and almost completely abolished increased *Ppargc1a* expression stimulated with AdipoRon in C2C12 myotubes (Fig. 1h), consistent with the report that increased PGC-1α expression mediated by adiponectin is dependent on Ca^{2+} signalling[21].

By using surface plasmon resonance, AdipoRon bound to both AdipoR1 and AdipoR2 (dissociation constant (K_d) of 1.8 and 3.1 μM; R_{max} of 14.6 and 8.6 resonance units (RU), respectively) in a saturable manner (Fig. 1j, k). We also performed radioactive binding and Scatchard analysis and verified the specific binding of AdipoRon to AdipoR1 and AdipoR2 (Extended Data Fig. 4).

Intravenous injection of AdipoRon (50 mg kg^{-1} body weight) significantly induced phosphorylation of AMPK in skeletal muscle and liver of wild-type mice but not *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout mice (Fig. 1l, m), indicating that AdipoRon could activate AMPK in skeletal muscle and liver via AdipoR1 and AdipoR2.

AdipoRon ameliorates diabetes via AdipoR

To clarify whether orally administered small-molecule AdipoR agonist AdipoRon would exhibit a pharmacokinetic profile suitable for *in vivo* evaluation in the mouse, we measured plasma concentrations of AdipoRon in C57BL/6 wild-type mice after oral administration of 50 mg kg^{-1} of AdipoRon, and found that the maximal concentration (C_{max}) of AdipoRon was 11.8 μM (Fig. 2a and Extended Data Fig. 5a).

To test the therapeutic potential of a small-molecule AdipoR agonist to treat insulin resistance and diabetes, the effects of orally administered AdipoRon were examined in high-fat-diet-induced obese mice. Oral administration of AdipoRon (50 mg kg^{-1} body weight) for 10 days did not significantly affect body weight (Fig. 2b) nor food intake (Fig. 2c) in mice on a high-fat diet, but it did significantly reduce fasting plasma glucose and insulin levels as well as glucose and insulin responses during oral glucose tolerance tests in wild-type mice treated with AdipoRon (Fig. 2d and Extended Data Fig. 5b, c). The decrease in glucose levels in the face of reduced plasma insulin levels indicates improved insulin sensitivity (Fig. 2d, f and Extended Data Fig. 5d, e). Notably, treatment of *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout mice with AdipoRon failed to ameliorate high-fat-diet-induced hyperglycaemia and hyperinsulinaemia (Fig. 2e, f and Extended Data Fig. 5f–i).

The glucose-lowering effect of exogenous insulin was also greater in AdipoRon-treated wild-type mice than in vehicle-treated control
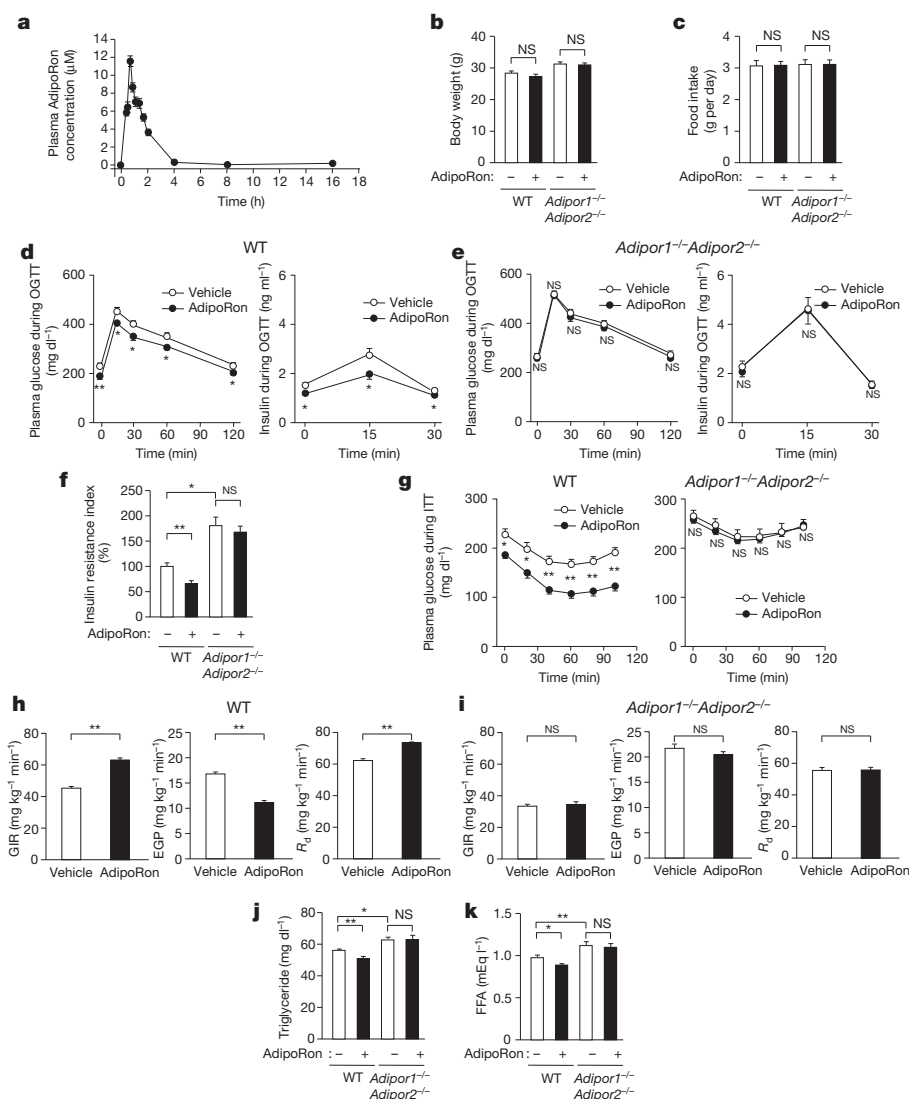
Figure 2 | AdipoRon improved insulin resistance, glucose intolerance and dyslipidaemia via AdipoR. a–g, Plasma AdipoRon concentrations (**a**), body weight (**b**), food intake (**c**), plasma glucose (**d**, **e**, **g**), plasma insulin (**d**, **e**) and insulin resistance index (**f**) during oral glucose tolerance test (OGTT) (1.0 g glucose per kg body weight) (**d**, **e**) or during insulin tolerance test (ITT) (0.5 U insulin per kg body weight) (**g**) in wild-type (WT) and $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice, treated with or without AdipoRon (50 mg per kg body weight). **h, i**, Glucose infusion rate (GIR), endogenous glucose production (EGP) and rates of glucose disposal (R_d) during hyperinsulinaemic euglycaemic clamp study in wild-type and $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice, treated with or without AdipoRon (50 mg per kg body weight). **j, k**, Plasma triglyceride (**j**) and free fatty acid (FFA) (**k**) in wild-type and $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice, treated with or without AdipoRon (50 mg per kg body weight). All values are presented as mean ± s.e.m. **a**, $n = 12–32$; **b–g, j, k**, $n = 10$ each; **h, i**, $n = 5$ each; *$P < 0.05$ and **$P < 0.01$ compared to control or as indicated. NS, not significant.

wild-type mice (Fig. 2g, left, and Extended Data Fig. 5j, k), which was not observed in $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice (Fig. 2g, right, and Extended Data Fig. 5l, m).

We examined whether a similar chemical analogue of AdipoRon that could activate AMPK via AdipoR would have an antidiabetic effect. Consistent with this, we observed that another similar chemical analogue of AdipoRon, compound 112254 (Extended Data Fig. 6a), could activate AMPK (Fig. 1c) and at the same time ameliorate both glucose intolerance and insulin resistance (Extended Data Fig. 6c–f). Conversely, we observed that another compound, 165073 (Extended Data Fig. 6b), could not activate AMPK (Fig. 1c), ameliorate glucose intolerance, nor ameliorate insulin resistance (Extended Data Fig. 6g–j).

We performed hyperinsulinaemic euglycaemic clamps in mice on a high-fat diet after 10 days of treatment. The glucose infusion rate was significantly increased (Fig. 2h, left), the endogenous glucose production was significantly suppressed (Fig. 2h, middle), and the glucose disposal rate was significantly increased (Fig. 2h, right) in AdipoRon-treated wild-type mice. None of these parameters was improved on AdipoRon treatment in $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice (Fig. 2i).

We next examined the effects of AdipoRon on lipid metabolism. Treatment with AdipoRon for 10 days reduced plasma concentrations of triglycerides and free fatty acid (FFA) in wild-type mice fed a high-fat diet (Fig. 2j, k), an effect that was not observed in $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice (Fig. 2j, k).

AdipoRon activates AdipoR1–AMPK–PGC-1α pathways

In skeletal muscle of wild-type mice, AdipoRon increased the expression of genes involved in mitochondrial biogenesis such as *Ppargc1a* and oestrogen-related receptor-α (*Esrra*)[29], mitochondrial DNA replication/translation such as mitochondrial transcription factor A (*Tfam*), and oxidative phosphorylation such as cytochrome *c* oxidase subunit II (*mt-Co2*) (Fig. 3a). AdipoRon also increased mitochondrial DNA content in the skeletal muscle of wild-type mice (Fig. 3b). These effects were completely obliterated in $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice (Fig. 3a, b).

AdipoRon increased the levels of oxidative, high endurance type I fibre[30] marker troponin I (slow) (*Tnni1*) in the skeletal muscle of wild-type mice (Fig. 3a) but not in $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice (Fig. 3a). We challenged mice fed a high-fat diet with involuntary physical exercise by treadmill running and then assessed muscle endurance. AdipoRon significantly increased exercise endurance in wild-type mice, but not in $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice (Fig. 3c) fed a high-fat diet.

We next examined the expression of metabolic genes and found that AdipoRon significantly increased the expression of genes involved in fatty-acid oxidation such as medium-chain acyl-CoA dehydrogenase (*Acadm*) (Fig. 3a), which was associated with decreased triglyceride content[31] (Extended Data Fig. 7a), in the skeletal muscle of wild-type mice but not of $Adipor1^{-/-}$ $Adipor2^{-/-}$ double-knockout mice fed a high-fat diet.
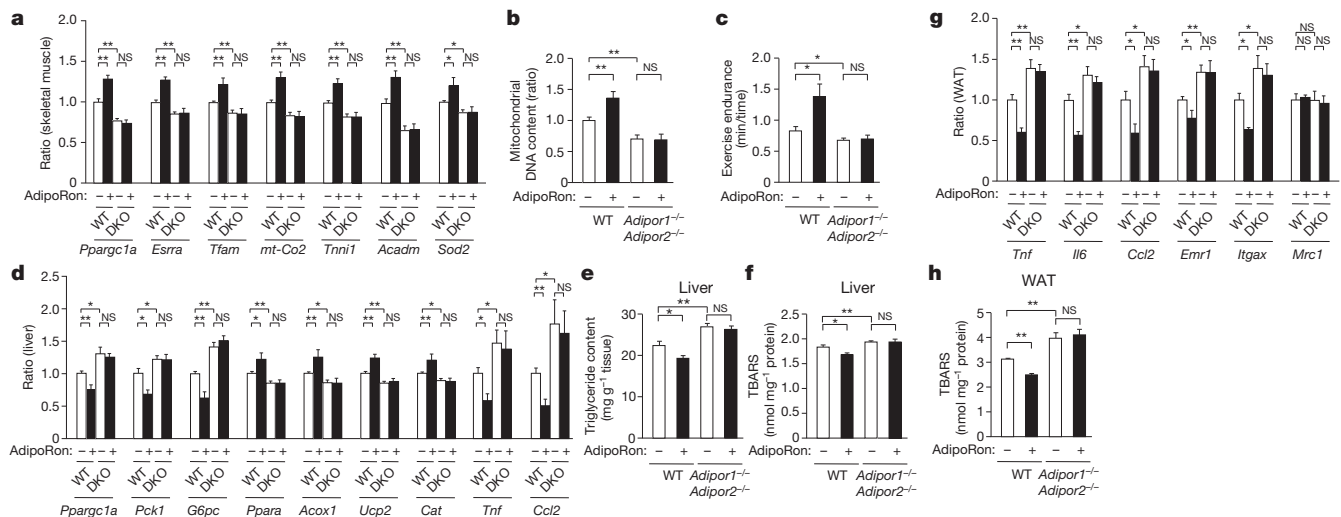
Figure 3 | AdipoRon increased mitochondria biogenesis in muscle, reduced tissue triglyceride content in liver and decreased oxidative stress and inflammation in liver and WAT. a–h, *Ppargc1a, Esrra, Tfam, mt-Co2, Tnni1, Acadm* and *Sod2* mRNA levels (**a**), mitochondrial content as assessed by mitochondrial DNA copy number (**b**) in skeletal muscle, exercise endurance (**c**), *Ppargc1a, Pck1, G6pc, Ppara, Acox1, Ucp2, Cat, Tnf* and *Ccl2* mRNA levels

(**d**), tissue triglyceride content (**e**), TBARS (**f**) in liver and *Tnf, Il6, Ccl2, Emr1, Itgax* and *Mrc1* mRNA levels (**g**) and TBARS (**h**) in WAT, from wild-type and *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout (DKO) mice treated with or without AdipoRon (50 mg per kg body weight). All values are presented as mean ± s.e.m. **a, b, d–h,** $n = 10$ each; **c,** $n = 5$ each; *$P < 0.05$ and **$P < 0.01$ compared to control or as indicated. NS, not significant.

AdipoRon significantly increased the expression levels for oxidative stress-detoxifying genes such as manganese superoxide dismutase (*Sod2*) (Fig. 3a), and decreased oxidative stress markers[32] such as thiobarbituric acid reactive substance (TBARS) (Extended Data Fig. 7b), in the skeletal muscle of wild-type mice but not of *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice fed a high-fat diet.

AdipoRon also activates AdipoR2–PPAR-α pathways

We examined whether AdipoRon could activate AdipoR1 and AdipoR2 pathways in the liver. The activation of AdipoR1–AMPK pathway in the liver has been reported to reduce the expression of genes involved in hepatic gluconeogenesis such as *Ppargc1a*, phosphoenolpyruvate carboxykinase 1 (*Pck1*)[20,33] and glucose-6-phosphatase (*G6pc*). As predicted by these earlier studies, we found that AdipoRon significantly decreased the expression of *Ppargc1a, Pck1* and *G6pc* in the liver of wild-type (Fig. 3d) but not of *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice (Fig. 3d) fed a high-fat diet.

Activation of AdipoR2 can increase PPAR-α levels and activate PPAR-α pathways, leading to increased fatty-acid oxidation and reduction of oxidative stress[20]. AdipoRon increased the expression levels of the gene encoding PPAR-α itself (*Ppara*) and its target genes[16], including genes involved in fatty-acid combustion such as acyl-CoA oxidase (*Acox1*), genes involved in energy dissipation such as uncoupling protein 2 (*Ucp2*), and genes encoding oxidative stress detoxifying enzymes such as catalase (*Cat*) in the liver of wild-type (Fig. 3d) but not of *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice (Fig. 3d) fed a high-fat diet. AdipoRon significantly reduced triglyceride content (Fig. 3e) and oxidative stress[32], as measured by TBARS (Fig. 3f), in the liver of wild-type mice but not of *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice (Fig. 3e, f) fed a high-fat diet.

Notably, orally administered AdipoRon reduced the expression levels of the genes encoding pro-inflammatory cytokines such as TNF-α (*Tnf*)[34] and MCP-1 (*Ccl2*) in the liver of wild-type mice (Fig. 3d) but not of *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice (Fig. 3d) fed a high-fat diet.

AdipoRon decreases inflammation

AdipoRon reduced the expression levels of genes encoding pro-inflammatory cytokines[35–37] such as *Tnf*, IL-6 (*Il6*) and *Ccl2* in the white

adipose tissue (WAT) of wild-type mice but not of *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice fed a high-fat diet (Fig. 3g). Notably, AdipoRon reduced TBARS (Fig. 3h) and reduced levels of macrophage markers such as F4/80 (*Emr1*), and especially the levels of markers for classically activated M1 macrophages such as CD11c (*Itgax*)[38]—but not the levels of markers for the alternatively activated M2 macrophages such as CD206 (*Mrc1*)—in the WAT of wild-type mice fed a high-fat diet (Fig. 3g), whereas these changes were not observed in *Adipor1$^{-/-}$ Adipor2$^{-/-}$* double-knockout mice (Fig. 3g, h).

AdipoRon ameliorates diabetes in *db/db* mice

We next studied the effects of AdipoRon (50 mg kg^{-1} body weight) in a genetically obese rodent model (*Lepr$^{-/-}$* (also known as *db/db*) mice); *db/db* mice fed a normal chow diet exhibit decreased plasma adiponectin concentrations[6,10]. As was expected[13], intraperitoneal injection of adiponectin into *db/db* mice reduced plasma glucose levels (Fig. 4a, left and right panels). Interestingly, orally administered AdipoRon also significantly reduced plasma glucose levels as quickly and potently as did intraperitoneal adiponectin injection in *db/db* mice (Fig. 4a, middle and right panels).

Without affecting body weight, food intake, liver weight and WAT weight (Fig. 4b–e), orally administered AdipoRon for 2 weeks significantly ameliorated glucose intolerance, insulin resistance and dyslipidaemia in *db/db* mice fed a normal chow diet (Fig. 4f–i).

In the skeletal muscle of *db/db* mice fed a normal chow diet, AdipoRon significantly increased the expression levels of genes involved in mitochondrial biogenesis functions and DNA content (Fig. 5a, b), and also *Acadm* and *Sod2* (Fig. 5a), which were associated with decreased triglyceride content and TBARS (Fig. 5c, d), respectively. In the liver, AdipoRon significantly decreased the expression of *Ppargc1a, Pck1* and *G6pc* (Fig. 5e), increased the expression of *Ppara* and its target genes (Fig. 5e). Therefore, Adipron significantly reduced triglyceride content (Fig. 5f), oxidative stress (Fig. 5g) and reduced the expression levels of genes encoding pro-inflammatory cytokines (Fig. 5e). In the WAT, AdipoRon reduced the expression levels of genes encoding pro-inflammatory cytokines and macrophage markers, especially the levels of markers for classically activated M1 macrophages, but not the levels of markers for the alternatively activated M2 macrophages (Fig. 5h).

Figure 4 | AdipoRon ameliorated insulin resistance, diabetes and dyslipidaemia in *db/db* mice.
a, Plasma glucose levels after intraperitoneal injection of adiponectin (30 μg per 10 g body weight) (left) or after oral administration of AdipoRon (50 mg per kg body weight) (middle). The area under the curve (AUC) of left and middle panels is shown on the right. **b–i**, Body weight (**b**), food intake (**c**), liver weight (**d**), WAT weight (**e**), plasma glucose (**f**, left, **g**), plasma insulin (**f**, middle) and insulin resistance index (**f**, right) during oral glucose tolerance test (OGTT) (1.0 g glucose per kg body weight) (**f**) or during insulin tolerance test (ITT) (0.75 U insulin per kg body weight) (**g**), plasma triglyceride (**h**) and free fatty acid (FFA) (**i**) in *db/db* mice under normal chow conditions, treated with or without AdipoRon (50 mg per kg body weight). All values are presented as mean ± s.e.m.
a, $n = 6$-7; **b–i**, $n = 10$ each from 2–3 independent experiments, $*P < 0.05$ and $**P < 0.01$ compared to control or as indicated. NS, not significant.

AdipoRon prolonged the shortened lifespan

Notably, *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout mice showed a shortened lifespan as compared with wild-type mice under both normal chow diet and high-fat diet conditions (Fig. 6a, b). Because a high-fat diet has been reported to shorten lifespan[39], we examined whether orally administered AdipoR agonists could prolong the shortened lifespan on a high-fat diet. Lifespan of *db/db* mice on a high-fat diet was markedly shortened as compared with that on a normal chow diet. Surprisingly, AdipoRon significantly rescued the shortened lifespan of *db/db* mice on a high-fat diet (Fig. 6c).

The decreased effects of adiponectin in obesity have been reported to have causal roles in the development of obesity-related diseases such as diabetes[40] and cardiovascular diseases[41]. There are two strategies to reverse reduced adiponectin effects. One is to increase the levels of adiponectin itself, such as through the injection of adiponectin. However, there are many difficulties associated with adiponectin injection, such as very high plasma concentrations of adiponectin and high-molecular-weight adiponectin multimers as highest activity form[42].

An alternative strategy is to activate adiponectin receptors. Both AdipoR1 and AdipoR2 have roles in the regulation of glucose and lipid metabolism, inflammation, and oxidative stress *in vivo*[20]. Therefore, the development of orally active small-molecule agonists for both AdipoR1 and AdipoR2 has long been sought. Here, we have identified and characterized an orally active synthetic small molecule that binds to and activates AdipoR1 and AdipoR2. So far, the top four hits obtained through the screening campaign have common structural motifs (Extended Data Fig. 8) (see additional results and discussion in Supplementary Information).

One of these small molecules, AdipoRon, binds to both AdipoR1 and AdipoR2 *in vitro* (K_d 1.8 and 3.1 μM; R_{max} 14.6 and 8.6 RU, respectively), activates AMPK, and increases PGC-1α levels and mitochondrial DNA content in myotubes (Fig. 1). When AdipoRon was administered orally to mice (50 mg per kg body weight), it was confirmed that the concentrations of AdipoRon in plasma (C_{max} of 11.8 μM) reached levels greater than the K_d values (AdipoR1, 1.8 μM; AdipoR2, 3.1 μM) (Fig. 2a). After the concentration reached the maximum as shown in Fig. 2a, the effect reached the maximum (Extended Data Fig. 5n), and the effect lasted for at least 8 h. Orally administered AdipoRon ameliorated insulin resistance, glucose intolerance and dyslipidaemia in mice fed a high-fat diet (Fig. 2d–k). Notably, these beneficial effects were completely obliterated in *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout mice (Fig. 2d–k) but partially preserved in *Adipor1*$^{-/-}$ or *Adipor2*$^{-/-}$ single-knockout mice (Extended Data Fig. 7c-g), indicating that AdipoRon works through both AdipoR1 and AdipoR2 *in vivo*.

Adiponectin ameliorated insulin resistance and glucose intolerance via multiple mechanisms including activation of AMPK, decreased oxidative stress, decreased tissue triglyceride content and suppression of inflammation[13,14]. AdipoRon exerted multiple effects very similar to those of adiponectin described above *in vivo*, and ameliorated insulin resistance and glucose intolerance via AdipoR1 and AdipoR2 in obese diabetic mice on a high-fat diet (Fig. 3).

In this study, we show that in skeletal muscle of obese diabetic mice such as wild-type mice on a high-fat diet (Fig. 3) and *db/db* mice (Figs 4 and 5), AdipoR1 and AdipoR2 agonists such as AdipoRon increase mitochondrial biogenesis, which was associated with increased
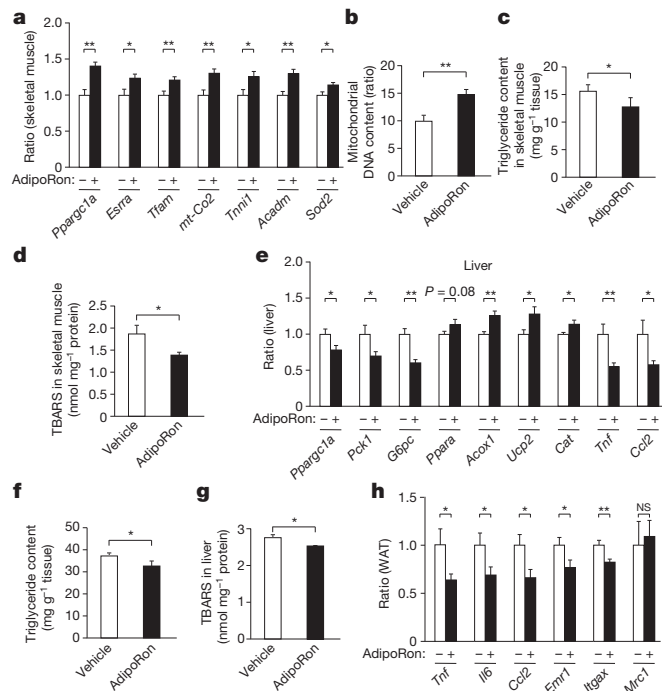
Figure 5 | **AdipoRon increased mitochondria biogenesis in muscle, reduced tissue triglyceride content and oxidative stress in muscle and liver, and decreased inflammation in liver and WAT of *db/db* mice.** **a–h**, *Ppargc1a*, *Esrra*, *Tfam*, *mt-Co2*, *Tnni1*, *Acadm* and *Sod2* mRNA levels (**a**), and mitochondrial content as assessed by mitochondrial DNA copy number (**b**), tissue triglyceride content (**c**) and TBARS (**d**) in skeletal muscle, *Ppargc1a*, *Pck1*, *G6pc*, *Ppara*, *Acox1*, *Ucp2*, *Cat*, *Tnf* and *Ccl2* mRNA levels (**e**), tissue triglyceride content (**f**) and TBARS (**g**) in liver, and *Tnf*, *Il6*, *Ccl2*, *Emr1*, *Itgax* and *Mrc1* mRNA levels (**h**) in WAT from *db/db* mice on a normal chow diet, treated with or without AdipoRon (50 mg per kg body weight). All values are presented as mean ± s.e.m. $n = 10$, *$P < 0.05$ and **$P < 0.01$ compared to control or as indicated. NS, not significant.

exercise endurance, and at the same time increase expression levels of genes involved in fatty-acid combustion, oxidative phosphorylation and reduction of oxidative stress (Figs 3, 5 and 6d). In liver, AdipoRon suppresses the expression of genes involved in gluconeogenesis, increases expression of PPAR-α target genes involved in fatty-acid combustion, and reduces oxidative stress (Figs 3, 5 and 6d). In WAT, AdipoRon reduces oxidative stress and pro-inflammatory cytokines, and the accumulation of M1 macrophages (Figs 3, 5 and 6d). Importantly, these effects resulted in reduced tissue triglyceride content in liver and muscle, and oxidative stress in liver, muscle and WAT, and decreased inflammation in liver and WAT (Figs 3–5 and 6d). These alterations collectively result in increased insulin sensitivity and glucose tolerance (Fig. 6d).

Therefore, we could expect AdipoRon to exert most, if not all, of the effects exerted by adiponectin, such as increased insulin sensitivity and glucose tolerance, as well as suppression of cardiovascular diseases and cancer, as previously reported[17,41,43]. Indeed, AdipoRon did prolong the shortened lifespan of obese diabetic mice (Fig. 6a–d).

Taken together, our findings show that the orally active small-molecule AdipoR agonist AdipoRon shifts the physiology of mice fed excess calorie towards that of mice fed a standard diet, modulates known longevity pathways, and improves health and prolongs lifespan. This study provides evidence that an orally available synthetic small-molecule AdipoR agonist at doses achievable *in vivo* can safely reduce many of the unhealthy and undesirable consequences of excess calorie intake and sedentary lifestyle, with an overall improvement in health and even lifespan, much like calorie restriction and exercise. Because virtually all current therapeutic modalities of type 2 diabetes require stringent adherence to diet and exercise and are associated with adverse effects such as hypoglycaemia and weight gain, AdipoRon provides a novel pre-emptive medicine and treatment modality. Orally active AdipoR agonists are a promising novel therapeutic approach for treating obesity-related disorders such as type 2 diabetes.

Figure 6 | **AdipoRon increased insulin sensitivity and glucose tolerance, and at the same time contributed to longevity of obese diabetic mice.** **a–c**, Kaplan–Meier survival curves for wild-type, *Adipor1*[−/−], *Adipor2*[−/−] and *Adipor1*[−/−] *Adipor2*[−/−] knockout mice on a normal chow diet (**a**) ($n = 50, 32$, 29 and 35, respectively) or high-fat diet (**b**) ($n = 47, 33, 35$ and 31, respectively), or for *db/db* mice treated with or without AdipoRon (30 mg per kg body weight) on a normal chow or high-fat diet ($n = 20$ each) (**c**). *P* values were derived from log-rank calculations. **d**, Scheme illustrating the mechanisms by which AdipoR1 and AdipoR2 agonist increases insulin sensitivity and glucose tolerance, and at the same time lifespan. (See also main text.)

METHODS SUMMARY

Mice. Mice were 6–10 weeks of age at the time of the experiment. The animal care and use procedures were approved by the Animal Care Committee of the University of Tokyo (see additional Methods in Supplementary Information).

Studies with C2C12 cells. Induction of myogenic differentiation was carried out according to a method described previously[21]. By day 5, the cells had differentiated into multinucleated contracting myotubes. C2C12 myotubes were used after myogenic differentiation in all experiments.

Survival. The wild-type, $Adipor1^{-/-}$, $Adipor2^{-/-}$, $Adipor1^{-/-}$ $Adipor2^{-/-}$ knockout mice and the db/db mice were maintained with food and water ad libitum. In these experiments, we used standard chow diet (CE-2, CLEA Japan Inc.) or high-fat diet 32 (CLEA Japan Inc.)[20]. For the experiment shown in Fig. 6a, b, wild-type ($n = 50$), $Adipor1^{-/-}$ ($n = 32$), $Adipor2^{-/-}$ ($n = 29$) and $Adipor1^{-/-}$ $Adipor2^{-/-}$ ($n = 35$) knockout mice fed a normal chow diet were used. For the experiment shown in Fig. 6b, wild-type ($n = 47$), $Adipor1^{-/-}$ ($n = 33$), $Adipor2^{-/-}$ ($n = 35$) and $Adipor1^{-/-}$ $Adipor2^{-/-}$ ($n = 31$) knockout mice on a high-fat diet were used. For the experiment shown in Fig. 6c, the db/db mice were randomly divided into three groups: a normal chow group (normal chow, $n = 20$), high-fat group (high fat, $n = 20$) and high-fat plus AdipoRon group (high fat + AdipoRon, $n = 20$), which were treated with AdipoRon at a daily dose of 30 mg kg^{-1} body weight. The survival rate was recorded daily. Survival curves were plotted using the Kaplan–Meier method.

Statistical analysis. Results are expressed as mean ± s.e.m. Differences between two groups were assessed using unpaired two-tailed t-tests. Data involving more than two groups were assessed by analysis of variance (ANOVA).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.
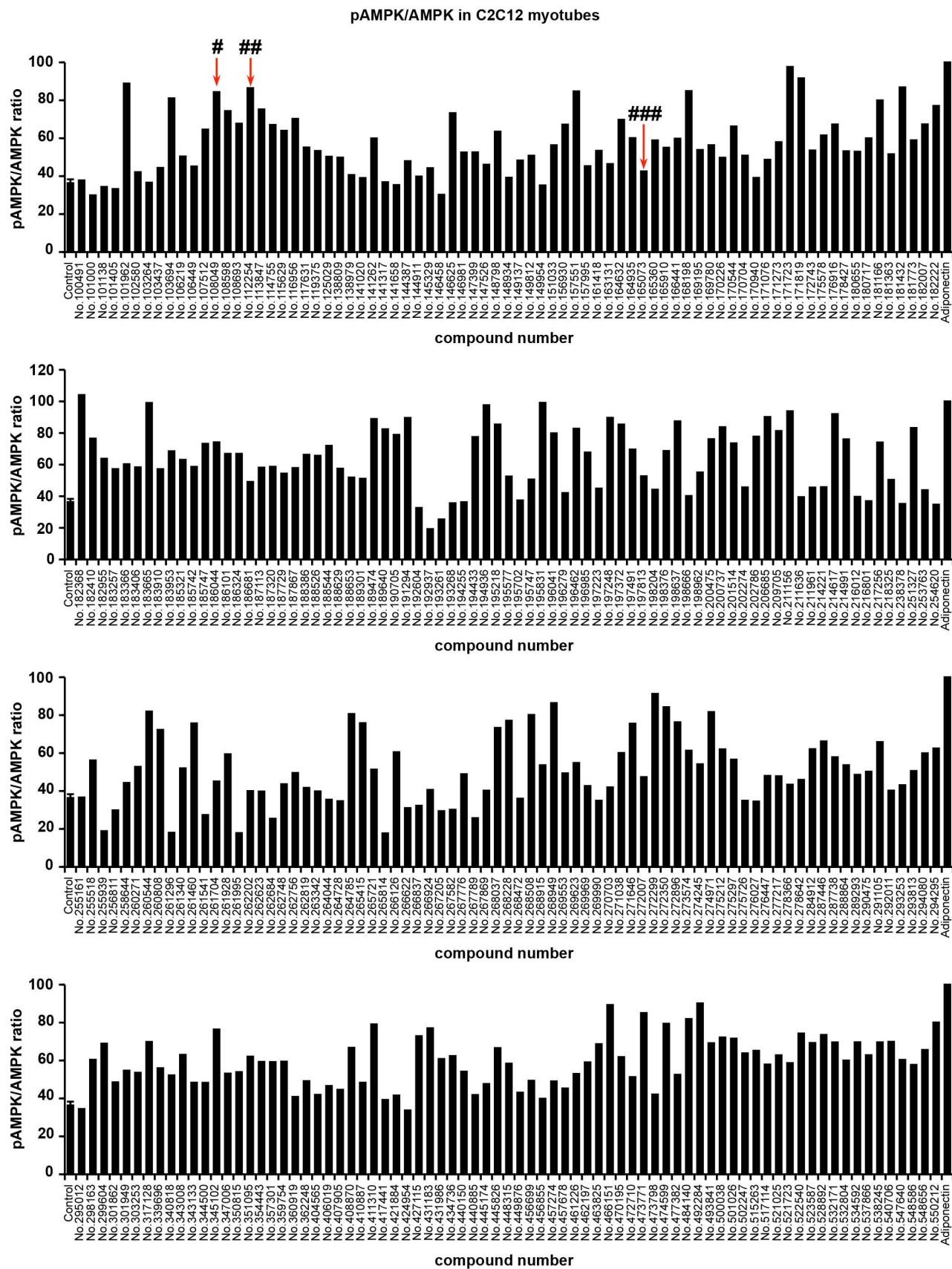
1. Gesta, S., Tseng, Y. H. & Kahn, C. R. Developmental origin of fat: tracking obesity to its source. *Cell* **131**, 242–256 (2007).
2. Olefsky, J. M. & Glass, C. K. Macrophages, inflammation, and insulin resistance. *Annu. Rev. Physiol.* **72**, 219–246 (2010).
3. Osler, M. E. & Zierath, J. R. Adenosine 5′-monophosphate-activated protein kinase regulation of fatty acid oxidation in skeletal muscle. *Endocrinology* **149**, 935–941 (2008).
4. LeRoith, D. & Accili, D. Mechanisms of disease: using genetically altered mice to study concepts of type 2 diabetes. *Nature Clin. Pract. Endocrinol. Metab.* **4**, 164–172 (2008).
5. Scherer, P. E., Williams, S., Fogliano, M., Baldini, G. & Lodish, H. F. A novel serum protein similar to C1q, produced exclusively in adipocytes. *J. Biol. Chem.* **270**, 26746–26749 (1995).
6. Hu, E., Liang, P. & Spiegelman, B. M. AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J. Biol. Chem.* **271**, 10697–10703 (1996).
7. Maeda, K. *et al.* cDNA cloning and expression of a novel adipose specific collagen-like factor, apM1 (AdiPose most abundant gene transcript 1). *Biochem. Biophys. Res. Commun.* **221**, 286–289 (1996).
8. Nakano, Y., Tobe, T., Choi-Miura, N. H., Mazda, T. & Tomita, M. Isolation and characterization of GBP28, a novel gelatin-binding protein purified from human plasma. *J. Biochem.* **120**, 803–812 (1996).
9. Hotta, K. *et al.* Plasma concentrations of a novel, adipose-specific protein, adiponectin, in type 2 diabetic patients. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1595–1599 (2000).
10. Yamauchi, T. *et al.* The fat-derived hormone adiponectin reverses insulin resistance associated with both lipoatrophy and obesity. *Nature Med.* **7**, 941–946 (2001).
11. Berg, A. H., Combs, T. P., Du, X., Brownlee, M. & Scherer, P. E. The adipocyte-secreted protein Acrp30 enhances hepatic insulin action. *Nature Med.* **7**, 947–953 (2001).
12. Fruebis, J. *et al.* Proteolytic cleavage product of 30-kDa adipocyte complement-related protein increases fatty acid oxidation in muscle and causes weight loss in mice. *Proc. Natl Acad. Sci. USA* **98**, 2005–2010 (2001).
13. Yamauchi, T. *et al.* Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nature Med.* **8**, 1288–1295 (2002).
14. Tomas, E. *et al.* Enhanced muscle fat oxidation and glucose transport by ACRP30 globular domain: acetyl-CoA carboxylase inhibition and AMP-activated protein kinase activation. *Proc. Natl Acad. Sci. USA* **99**, 16309–16313 (2002).
15. Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1**, 15–25 (2005).
16. Kersten, S., Desvergne, B. & Wahli, W. Roles of PPARs in health and disease. *Nature* **405**, 421–424 (2000).
17. Yamauchi, T. *et al.* Globular adiponectin protected ob/ob mice from diabetes and apoE deficient mice from atherosclerosis. *J. Biol. Chem.* **278**, 2461–2468 (2003).
18. Yamauchi, T. *et al.* Cloning of adiponectin receptors that mediate antidiabetic metabolic effects. *Nature* **423**, 762–769 (2003).
19. Wess, J. G-protein-coupled receptors: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *FASEB J.* **11**, 346–354 (1997).
20. Yamauchi, T. *et al.* Targeted disruption of AdipoR1 and AdipoR2 causes abrogation of adiponectin binding and metabolic actions. *Nature Med.* **13**, 332–339 (2007).
21. Iwabu, M. *et al.* Adiponectin and AdipoR1 regulate PGC-1α and mitochondria by Ca^{2+} and AMPK/SIRT1. *Nature* **464**, 1313–1319 (2010).
22. Richter, E. A. & Ruderman, N. B. AMPK and the biochemistry of exercise: implications for human health and disease. *Biochem. J.* **418**, 261–275 (2009).
23. Wu, Z. *et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115–124 (1999).
24. Handschin, C. & Spiegelman, B. M. The role of exercise and PGC1α in inflammation and chronic disease. *Nature* **454**, 463–469 (2008).
25. Cantó, C. *et al.* AMPK regulates energy expenditure by modulating NAD$^+$ metabolism and SIRT1 activity. *Nature* **458**, 1056–1060 (2009).
26. Paffenbarger, R. S. Jr *et al.* The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *N. Engl. J. Med.* **328**, 538–545 (1993).
27. Open Innovation Center for Drug Discovery. http://www.ocdd.u-tokyo.ac.jp/library_e.html (The University of Tokyo, 2012).
28. Hawley, S. A. *et al.* Characterization of the AMP-activated protein kinase kinase from rat liver and identification of threonine 172 as the major site at which it phosphorylates AMP-activated protein kinase. *J. Biol. Chem.* **271**, 27879–27887 (1996).
29. Mootha, V. K. *et al.* Errα and Gabpa/b specify PGC-1α-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. *Proc. Natl Acad. Sci. USA* **101**, 6570–6575 (2004).
30. Berchtold, M. W. *et al.* Calcium ion in skeletal muscle: its crucial role for muscle function, plasticity, and disease. *Physiol. Rev.* **80**, 1215–1265 (2000).
31. Shulman, G. I. Cellular mechanisms of insulin resistance. *J. Clin. Invest.* **106**, 171–176 (2000).
32. Brownlee, M. Biochemistry and molecular cell biology of diabetic complications. *Nature* **414**, 813–820 (2001).
33. Lochhead, P. A. *et al.* 5-aminoimidazole-4-carboxamide riboside mimics the effects of insulin on the expression of the 2 key gluconeogenic genes PEPCK and glucose-6-phosphatase. *Diabetes* **49**, 896–903 (2000).
34. Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science* **259**, 87–91 (1993).
35. Wellen, K. E. & Hotamisligil, G. S. Inflammation, stress, and diabetes. *J. Clin. Invest.* **115**, 1111–1119 (2005).
36. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).
37. Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112**, 1821–1830 (2003).
38. Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J. Clin. Invest.* **117**, 175–184 (2007).
39. Zhang, H. M. *et al.* Geldanamycin derivative ameliorates high fat diet-induced renal failure in diabetes. *PLoS ONE* **7**, e32746 (2012).
40. Li, S., Shin, H. J., Ding, E. L. & van Dam, R. M. Adiponectin levels and risk of type 2 diabetes: a systematic review and meta-analysis. *J. Am. Med. Assoc.* **302**, 179–188 (2009).
41. Pischon, T. *et al.* Plasma adiponectin levels and risk of myocardial infarction in men. *J. Am. Med. Assoc.* **291**, 1730–1737 (2004).
42. Pajvani, U. B. *et al.* Complex distribution, not absolute amount of adiponectin, correlates with thiazolidinedione-mediated improvement in insulin sensitivity. *J. Biol. Chem.* **279**, 12152–12162 (2004).
43. Luo, Z., Saha, A. K., Xiang, X. & Ruderman, N. B. AMPK, the metabolic syndrome and cancer. *Trends Pharmacol. Sci.* **26**, 69–76 (2005).

Extended Data Figure 1 | Phosphorylation of AMPK in C2C12 myotubes.
Phosphorylation of AMPK normalized to the amount of AMPK in C2C12
myotubes treated for 5 min with 15 μg ml^{-1} adiponectin or the indicated small-molecule compounds (10 μM). #, AdipoRon; ##, no. 112254; ###, no. 165073.
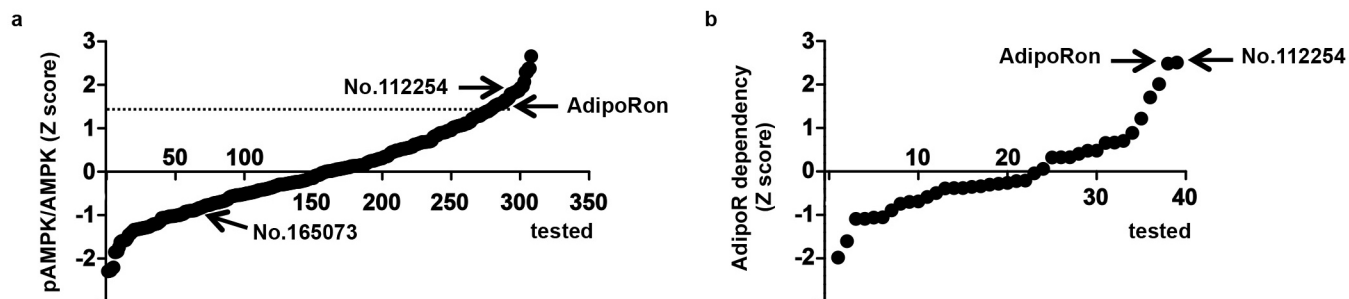
Extended Data Figure 2 | Distribution curves showing *Z* scores.
a, Distribution curve showing *Z* scores representing AMPK activity for all compounds tested in C2C12 myotubes shown in Extended Data Table 1 and Extended Data Fig. 1. The dashed line indicates the *Z* score cut-off for compounds scored as hits, which showed higher activity than 80% of that seen with adiponectin. **b**, Distribution curve showing *Z* scores representing AdipoR dependency of AMPK activation for 39 compounds tested in C2C12 myotubes shown in Extended Data Table 2. Indicated are the location of AdipoRon, another hit (no. 112254), and non-hit (no. 165073).

a

Complex I activity

b

***Adipor1* mRNA in C2C12 myotubes**

c

***Adipor2* mRNA in C2C12 myotubes**
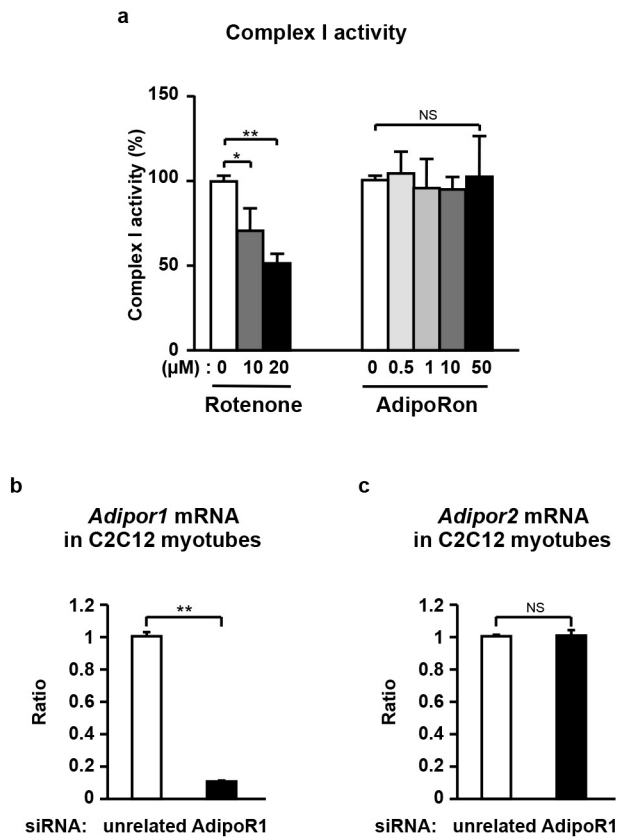
Extended Data Figure 3 | The effect of AdipoRon on complex I activity, and expression of *Adipor1* and *Adipor2* mRNA in C2C12 myotubes transfected with the indicated siRNA duplex. a, Complex I activities were measured with the indicated concentrations of rotenone or AdipoRon. **b, c,** *Adipor1* (**b**) and *Adipor2* (**c**) mRNA levels were analysed by RT–qPCR. All values are presented as mean ± s.e.m. **a**, $n = 3$–7; **b, c**, $n = 3$ each; *$P < 0.05$ and **$P < 0.01$ compared to control or unrelated siRNA cells. NS, not significant.
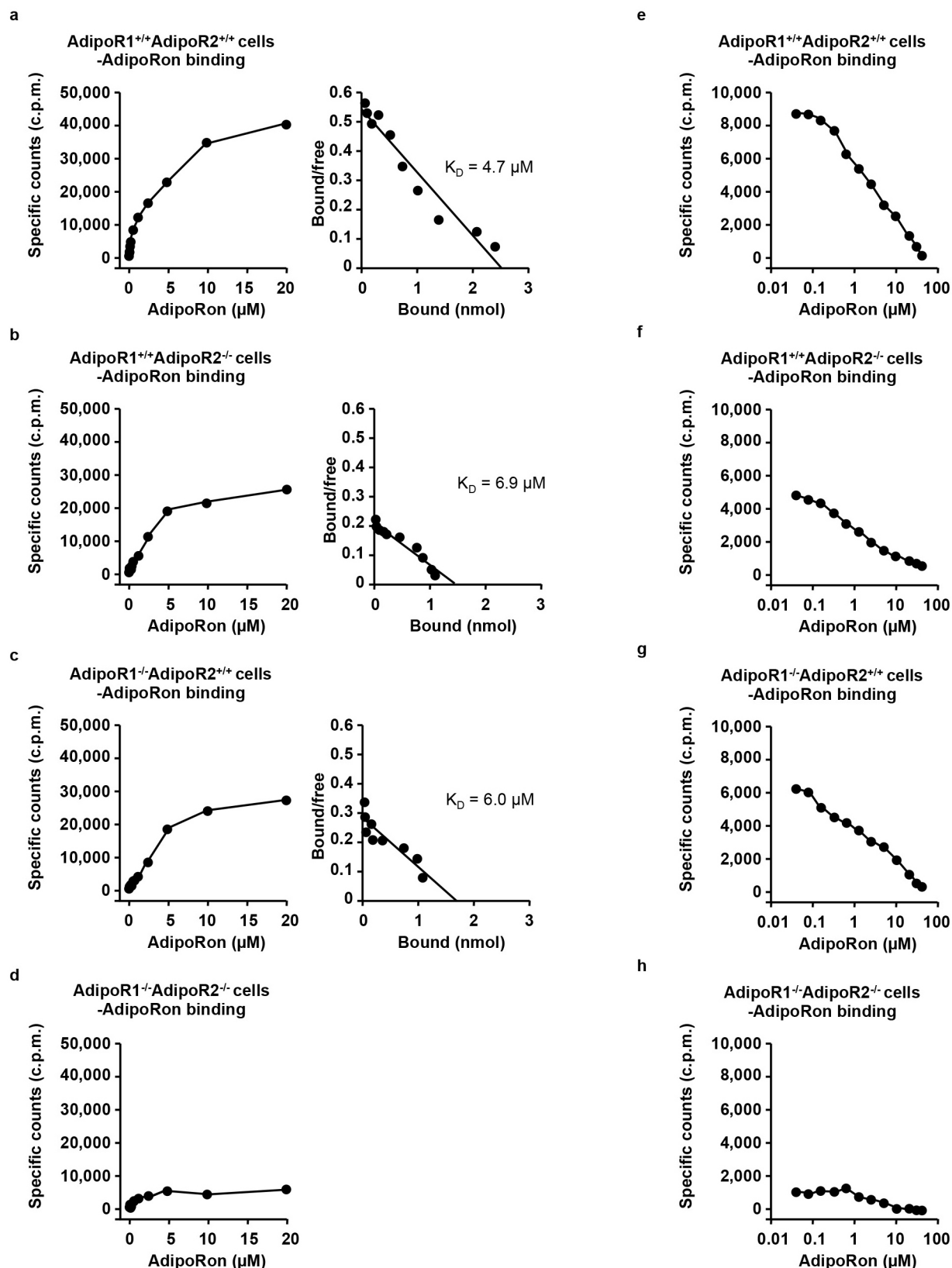
Extended Data Figure 4 | AdipoRon binding to AdipoR1 and AdipoR2.
a–d, Binding and Scatchard analyses of [^{3}H]AdipoRon to primary hepatocytes
from wild-type (**a**), *Adipor2*$^{-/-}$ knockout (**b**), *Adipor1*$^{-/-}$ knockout (**c**) and
Adipor1$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout (**d**) mice. **e–h**, Concentration-
dependent competitive [^{3}H]AdipoRon binding to primary hepatocytes from
wild-type (**e**), *Adipor2*$^{-/-}$ knockout (**f**), *Adipor1*$^{-/-}$ knockout (**g**) and
Adipor1$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout (**h**) mice. Binding analyses were
performed using the indicated concentrations of AdipoRon. c.p.m., counts per
minute.

a

raw data of Figure 2a

b WT
Vehicle

d WT
Vehicle

f *Adipor1⁻/⁻Adipor2⁻/⁻*
Vehicle

h *Adipor1⁻/⁻Adipor2⁻/⁻*
Vehicle

c WT
AdipoRon

e WT
AdipoRon

g *Adipor1⁻/⁻Adipor2⁻/⁻*
AdipoRon

i *Adipor1⁻/⁻Adipor2⁻/⁻*
AdipoRon

j WT
Vehicle

k WT
AdipoRon

l *Adipor1⁻/⁻Adipor2⁻/⁻*
Vehicle

m *Adipor1⁻/⁻Adipor2⁻/⁻*
AdipoRon

n Time course of glucose-lowering effect
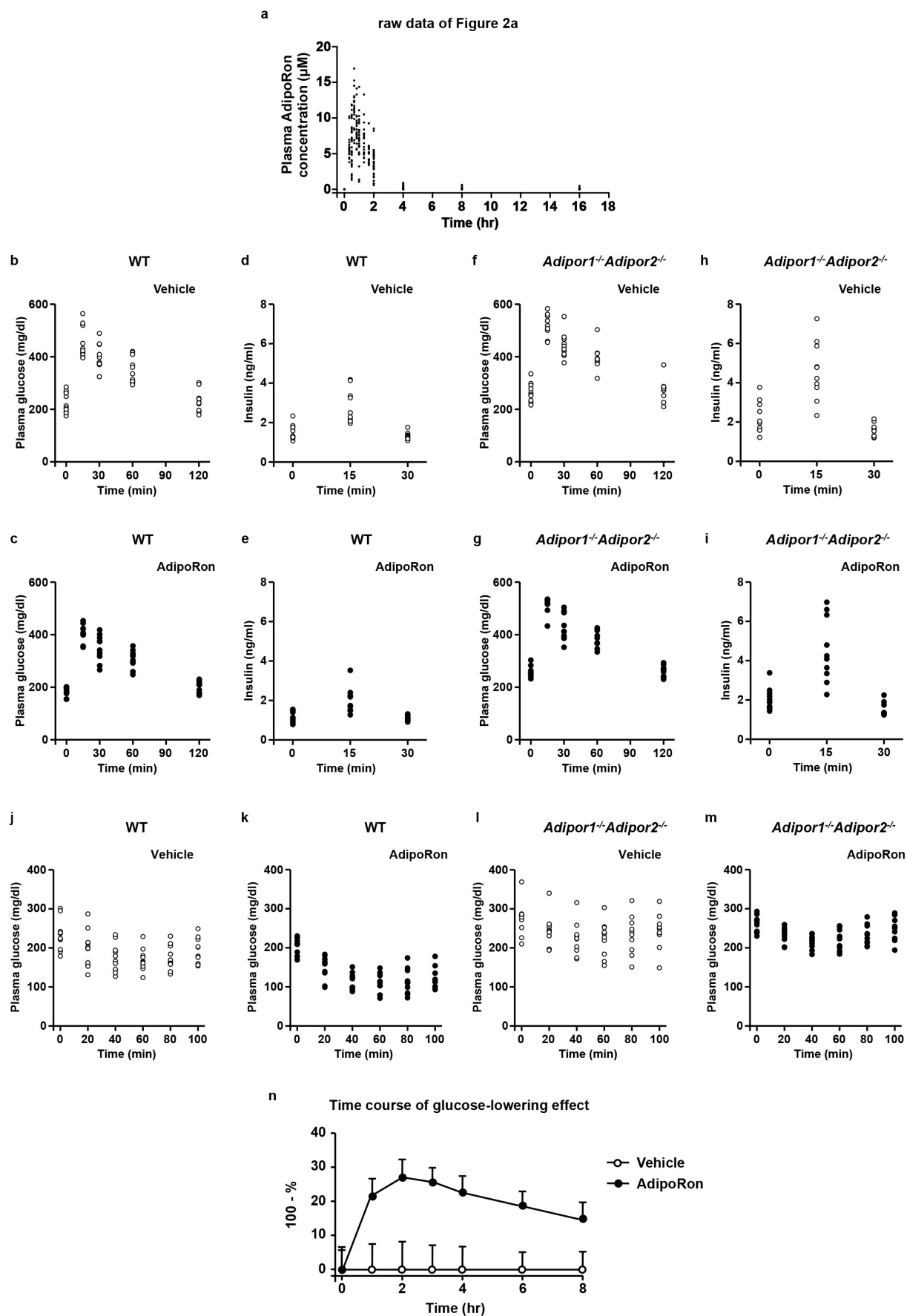
Extended Data Figure 5 | Raw data of Fig. 2 and time course of glucose-lowering effect of AdipoRon. a–m, Raw data of Fig. 2a (**a**), Fig. 2d, left (**b, c**), Fig. 2d, right (**d, e**), Fig. 2e, left (**f, g**), Fig. 2e, right (**h, i**), Fig. 2g, left (**j, k**) and Fig. 2g, right (**l, m**). **n**, Time course of glucose-lowering effect of AdipoRon. Data are calculated from data in Fig. 4a. The glucose-lowering effect of AdipoRon was obtained by the following equation and expressed as %: (vehicle plasma glucose − AdipoRon plasma glucose)/vehicle plasma glucose. All values are presented as mean ± s.e.m.

a

No.112254

b

No.165073

c WT

Plasma glucose levels during OGTT

○ Vehicle
● No.112254

p = 0.08
p = 0.1

Plasma insulin levels during OGTT

○ Vehicle
● No.112254

p = 0.05
p = 0.06

d *Adipor1⁻/⁻Adipor2⁻/⁻*

Plasma glucose levels during OGTT

○ Vehicle
● No.112254

NS
NS NS
NS

Plasma insulin levels during OGTT

○ Vehicle
● No.112254

NS
NS
NS NS

e Insulin resistance index

**
NS
**

No.112254: − + − +
WT   *Adipor1⁻/⁻ Adipor2⁻/⁻*

f Plasma glucose levels during ITT

WT

○ Vehicle
● No.112254

p = 0.08
**
*
p = 0.06
p = 0.08
p = 0.07

Adipor1⁻/⁻Adipor2⁻/⁻

NS NS NS NS NS NS

○ Vehicle
● No.112254

g WT

Plasma glucose levels during OGTT

○ Vehicle
● No.165073

NS
NS
NS
NS

Plasma insulin levels during OGTT

○ Vehicle
● No.165073

NS
NS
NS NS

h *Adipor1⁻/⁻Adipor2⁻/⁻*

Plasma glucose levels during OGTT

○ Vehicle
● No.165073

NS
NS
NS
NS

Plasma insulin levels during OGTT

○ Vehicle
● No.165073

NS
NS NS

i Insulin resistance index

**
NS
NS

No.165073 : − + − +
WT   *Adipor1⁻/⁻ Adipor2⁻/⁻*

j Plasma glucose levels during ITT

WT

○ Vehicle
● No.165073

NS NS NS NS NS NS

Adipor1⁻/⁻Adipor2⁻/⁻

NS NS NS NS NS NS

○ Vehicle
● No.165073

Extended Data Figure 6 | The effects of compounds 112254 and 165073 on insulin resistance and glucose intolerance via AdipoR. **a**, **b**, Chemical structures of compounds 112254 (**a**) and 165073 (**b**). **c–j**, Plasma glucose (**c** left, **d** left, **f**, **g** left, **h** left, **j**), plasma insulin (**c** right, **d** right, **g** right, **h** right) and insulin resistance index (**e**, **i**) during oral glucose tolerance test (OGTT) (1.0 g glucose per kg body weight) (**c**, **d**, **g**, **h**) or during insulin tolerance test (ITT) (0.5 U insulin per kg body weight) (**f**, **j**), in wild-type and $Adipor1^{-/-} Adipor2^{-/-}$ double-knockout mice, treated with or without compounds 112254 or 165073 (50 mg per kg body weight). All values are presented as mean ± s.e.m. **c–f**, $n = 10$ each; **g–j**, $n = 7$ each from 2, 3 independent experiments, *$P < 0.05$ and **$P < 0.01$ compared to control or as indicated. NS, not significant.

a

**TG
in skeletal muscle**

b

**TBARS
in skeletal muscle**

c

WT

d

Adipor1⁻/⁻

**Plasma glucose levels
during OGTT**

**Plasma insulin levels
during OGTT**

**Plasma glucose levels
during OGTT**

**Plasma insulin levels
during OGTT**

e

Adipor2⁻/⁻

f

Adipor1⁻/⁻ Adipor2⁻/⁻

g

Insulin resistance index
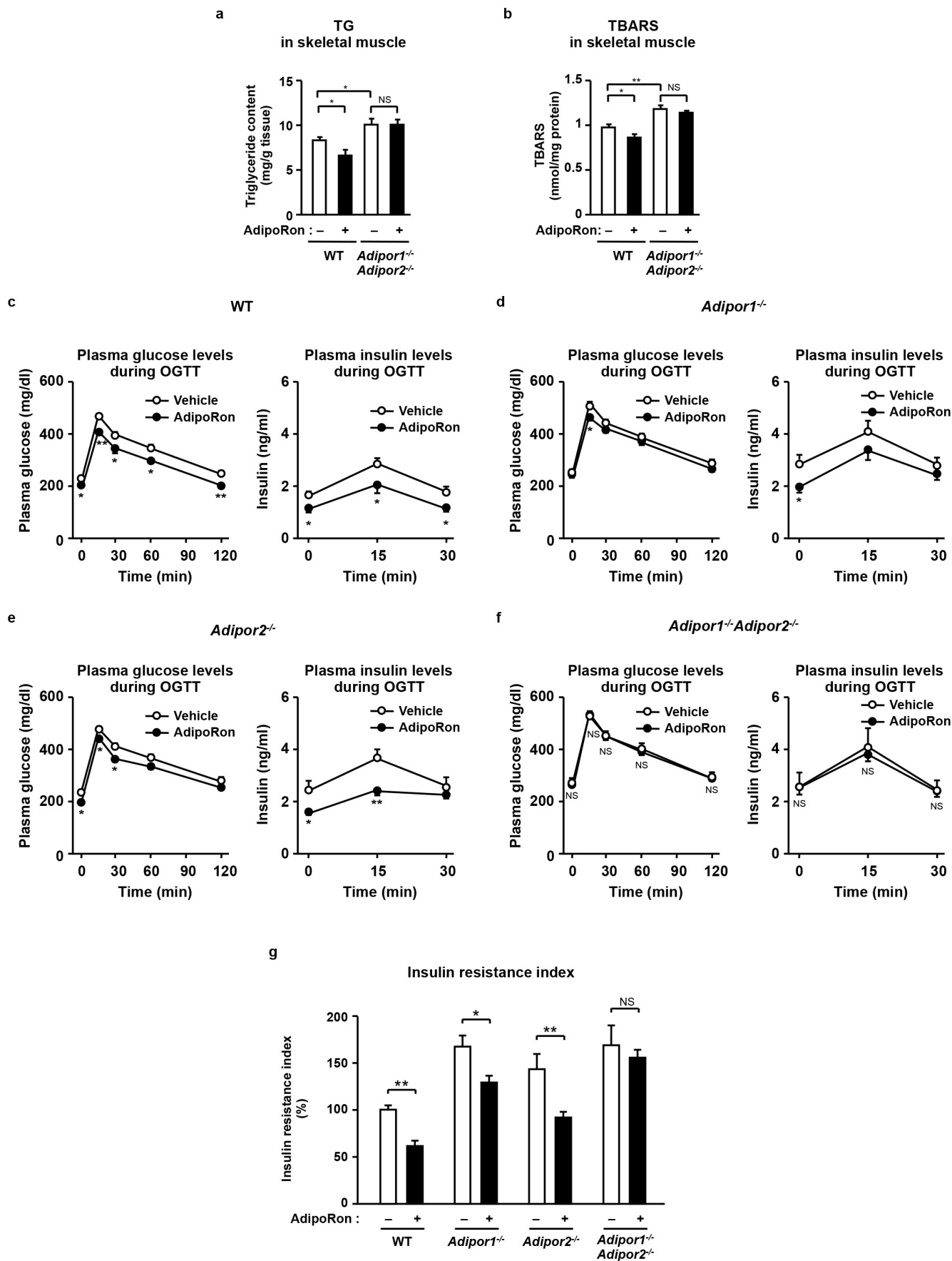
Extended Data Figure 7 | The effects of AdipoRon on glucose metabolism in *Adipor1*$^{-/-}$, *Adipor2*$^{-/-}$ and *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ mice. a, Triglyceride content (**a**) and TBARS (**b**) in skeletal muscle from wild-type or *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ double-knockout mice treated with or without AdipoRon (50 mg per kg body weight). **c–g**, The effects of AdipoRon on glucose metabolism in *Adipor1*$^{-/-}$, *Adipor2*$^{-/-}$ and *Adipor1*$^{-/-}$ *Adipor2*$^{-/-}$ mice.

Plasma glucose (**c–f**, left panels), plasma insulin (**c–f**, right panels) and insulin resistance index (**g**) during oral glucose tolerance test (OGTT) (1.0 g glucose per kg body weight). All values are presented as mean ± s.e.m. **a–d**, **f**, $n = 10$ each; **e**, $n = 7$ each; **g**, $n = 7$–10; *$P < 0.05$ and **$P < 0.01$ compared to vehicle mice. NS, not significant.

a AdipoRon

b No.168198

c No.112254

d No.103694

e pAMPK/AMPK

f AdipoR-dependency
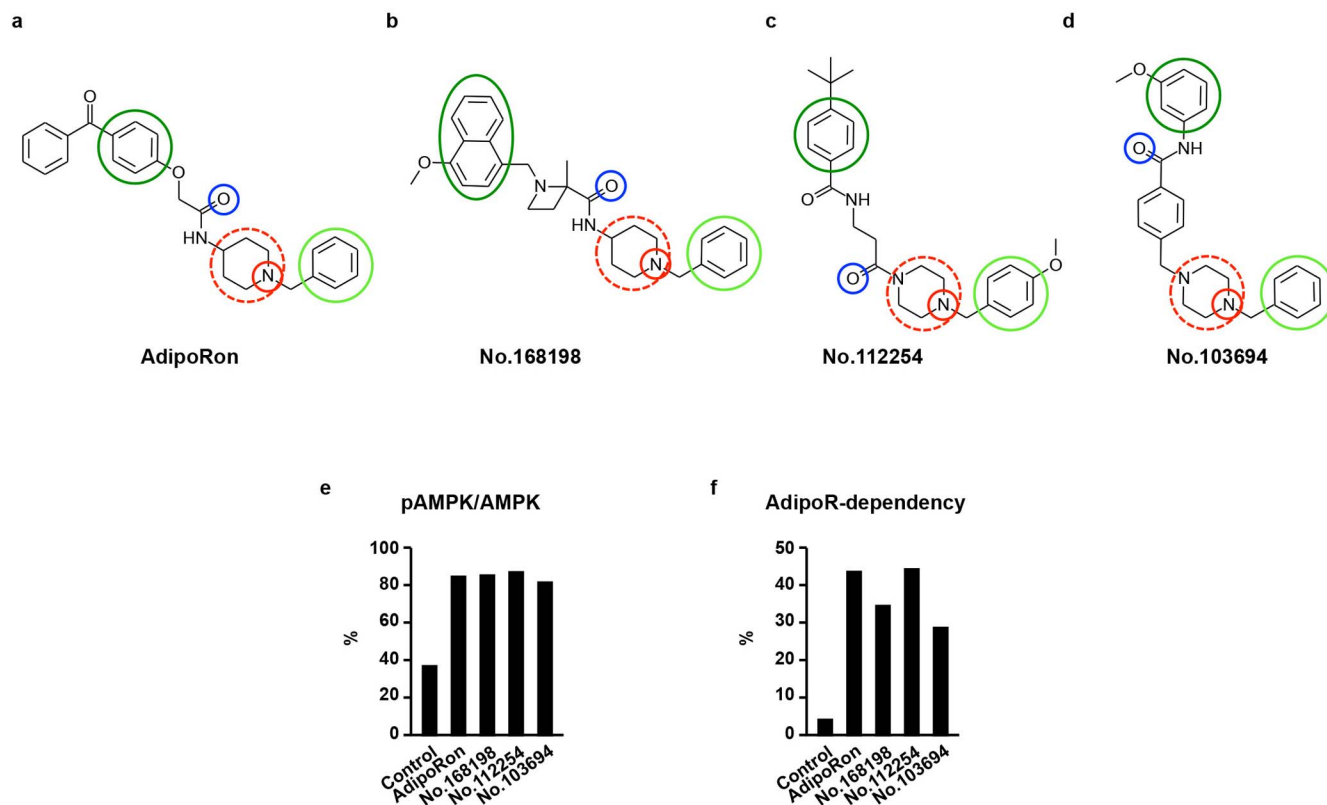
Extended Data Figure 8 | Chemical structures and AdipoR dependency of AMPK activation. **a–d**, Chemical structures of AdipoRon (**a**), compound 168198 (**b**), compound 112254 (**c**) and compound 103694 (**d**). Within the 1-benzyl 4-substituted 6-membered cyclic amine moiety, the cyclic amine moiety is surrounded by a dashed red circle, and the aromatic ring is surrounded by a light green circle. Cyan and dark green circles surround the carbonyl group and the terminal aromatic ring, respectively, located on the opposite side from the benzyl cyclic amine. **e**, Phosphorylation and amount of AMPK in C2C12 myotubes treated for 5 min with the indicated small-molecule compounds. Phosphorylation and amount of AMPK in C2C12 myotubes, treated for 5 min with the indicated small-molecule compounds (10 μM) (% relative to adiponectin). **f**, AdipoR dependency of AMPK activation. Phosphorylation and amount of AMPK in C2C12 myotubes and transfected with or without the AdipoR1 siRNA duplex, treated for 5 min with the indicated small molecule. AdipoR-dependency ratios were obtained by the following equation: 100 − (ratio for those transfected with the AdipoR1 siRNA duplex/ratio for those transfected without the AdipoR1 siRNA duplex) × 100 (%).

Extended Data Table 1 | Values of phosphorylation of AMPK in C2C12 myotubes

Compounds	%	Compounds	%	Compounds	%	Compounds	%	Compounds	%
Control	36.30	No.171076	48.67	No.197223	44.83	No.266924	40.60	No.354443	59.29
No.100491	37.78	No.171273	57.93	No.197248	89.63	No.267205	29.40	No.357301	59.19
No.101000	29.90	No.171723	97.60	No.197372	85.37	No.267582	30.14	No.359754	59.41
No.101138	34.34	No.171819	91.59	No.197491	69.57	No.267776	48.86	No.360919	40.83
No.101405	33.22	No.172743	53.55	No.197813	52.64	No.267789	25.70	No.362248	49.13
No.101962	88.84	No.175578	61.50	No.198204	44.15	No.267869	40.20	No.404565	41.91
No.102580	42.12	No.176916	67.27	No.198376	68.64	No.268037	73.29	No.406019	46.57
No.103264	36.58	No.178427	53.14	No.198637	87.39	No.268228	77.07	No.407905	44.57
No.103437	44.36	No.180655	52.89	No.198666	40.14	No.268472	35.97	No.408870	66.72
No.103694	81.05	No.180717	59.98	No.198962	55.06	No.268508	80.12	No.410887	48.29
No.106219	50.39	No.181166	79.96	No.200475	76.10	No.268915	53.61	No.411310	79.06
No.106449	45.13	No.181363	51.57	No.200737	83.71	No.268949	86.36	No.417441	39.23
No.107512	64.56	No.181432	86.82	No.201514	73.48	No.269553	49.33	No.421884	41.55
No.108049	84.33 #	No.181773	58.92	No.202274	45.61	No.269623	54.80	No.424954	33.68
No.108598	74.32	No.182007	67.31	No.202786	77.72	No.269969	42.64	No.427115	72.84
No.108693	67.68	No.182222	77.05	No.206685	90.15	No.269990	34.92	No.431183	77.01
No.112254	86.43 ##	No.182368	104.02	No.209705	81.28	No.270703	41.94	No.431986	60.80
No.113847	75.21	No.182410	76.46	No.211156	93.79	No.271038	60.07	No.434736	62.35
No.114755	67.02	No.182955	63.72	No.211636	39.43	No.271646	75.51	No.440150	54.14
No.115629	63.93	No.183257	57.18	No.211961	45.49	No.272007	47.33	No.440885	41.80
No.116956	70.23	No.183366	60.28	No.214221	45.73	No.272299	91.22	No.445174	47.63
No.117631	55.13	No.183406	58.30	No.214617	91.97	No.272350	84.20	No.445826	66.56
No.119375	53.30	No.183665	99.03	No.214991	76.02	No.272896	76.26	No.448315	58.34
No.125029	50.25	No.183910	57.14	No.216012	39.64	No.273574	61.28	No.449876	43.06
No.138809	49.79	No.183953	68.48	No.216801	36.89	No.274245	54.11	No.456699	49.34
No.138979	40.60	No.185321	63.03	No.217256	74.01	No.274971	81.64	No.456855	39.85
No.141020	39.04	No.185742	58.62	No.218325	50.36	No.275212	61.99	No.457274	49.00
No.141262	59.92	No.185747	73.20	No.238378	35.16	No.275297	56.55	No.457678	45.26
No.141317	36.80	No.186044	74.11	No.251327	83.18	No.275726	34.88	No.461226	52.94
No.141658	35.31	No.186101	66.82	No.253763	43.85	No.276027	34.47	No.462197	59.02
No.144387	47.92	No.186324	66.83	No.254620	34.68	No.276447	47.99	No.463825	68.63
No.144911	39.80	No.186681	49.05	No.255161	36.59	No.277217	47.80	No.466151	89.26
No.145329	44.22	No.187113	58.10	No.255518	56.12	No.278366	43.35	No.470195	61.81
No.146458	30.20	No.187320	58.64	No.255939	18.84	No.278642	45.88	No.472710	51.26
No.146625	73.23	No.187729	54.34	No.256811	29.84	No.284912	62.07	No.473771	85.03
No.146981	52.51	No.187867	57.86	No.258644	44.27	No.287446	66.18	No.473798	42.09
No.147399	52.57	No.188386	66.24	No.260271	52.74	No.287738	57.82	No.474599	79.39
No.147526	46.05	No.188526	65.61	No.260544	81.90	No.288864	53.61	No.477382	52.49
No.148798	63.49	No.188544	71.91	No.260808	72.25	No.289293	48.55	No.484140	81.88
No.148934	39.19	No.188629	57.46	No.261296	18.00	No.290475	50.20	No.492284	90.14
No.149137	48.26	No.188653	51.83	No.261340	51.95	No.291105	65.79	No.493841	69.13
No.149812	50.79	No.189301	51.08	No.261460	75.64	No.292011	40.21	No.500038	72.15
No.149954	35.12	No.189474	88.87	No.261541	27.34	No.293253	43.08	No.501026	71.56
No.151033	56.30	No.189640	82.40	No.261704	45.05	No.293813	50.55	No.502247	63.79
No.156930	67.18	No.190705	78.84	No.261928	59.40	No.294080	59.90	No.515263	65.13
No.157551	84.73	No.191294	89.56	No.261995	17.83	No.294295	62.45	No.517114	57.89
No.157995	45.27	No.192604	32.60	No.262202	39.97	No.295012	34.42	No.521025	62.76
No.161418	53.37	No.192937	19.22	No.262623	39.76	No.298163	60.39	No.521723	58.64
No.163131	46.39	No.193261	25.35	No.262684	25.39	No.299604	68.90	No.522540	74.24
No.164632	69.73	No.193268	35.56	No.262748	43.59	No.301862	48.50	No.523587	69.25
No.164935	60.08	No.194255	36.25	No.262756	49.53	No.301949	54.67	No.528892	73.49
No.165073	42.52 ###	No.194433	77.46	No.262819	41.67	No.303253	53.55	No.532171	69.57
No.165360	58.88	No.194936	97.60	No.263342	39.83	No.317128	69.80	No.532804	60.00
No.165910	55.01	No.195218	85.39	No.264044	35.43	No.339696	55.94	No.534592	69.62
No.166441	59.79	No.195577	52.49	No.264728	34.64	No.340818	52.17	No.537866	62.81
No.168198	84.85	No.195702	37.37	No.264785	80.62	No.343008	62.98	No.538245	69.60
No.169195	53.83	No.195747	50.57	No.265415	75.79	No.343133	48.31	No.540706	69.88
No.169780	56.32	No.195831	99.10	No.265721	51.29	No.344500	48.25	No.547640	60.33
No.170226	49.74	No.196041	79.83	No.265814	17.69	No.345102	76.36	No.548586	57.69
No.170544	66.23	No.196279	42.02	No.266126	60.45	No.347006	53.11	No.548656	65.56
No.170704	50.84	No.196462	82.72	No.266622	31.01	No.350815	53.89	No.550212	80.04
No.170940	39.11	No.196985	67.67	No.266837	32.24	No.351095	62.03	Adiponectin	100.00

Phosphorylation of AMPK normalized to the amount of AMPK in C2C12 myotubes treated for 5 min with 15 µg ml^{-1} adiponectin or the indicated small-molecule compounds (10 µM) (% relative to adiponectin).
#, AdipoRon; ##, no.112254; ###, no.165073.

Extended Data Table 2 | Phosphorylation of AMPK in AdipoR knock-down C2C12 myotubes

	pAMPK/AMPK (ratio)	
Compounds	unrelated siRNA	AdipoR1 siRNA
Control	1.00	0.96
No.101962	5.35	5.10
No.103694	4.67	3.34
No.108049	4.75	2.68
No.112254	3.70	2.07
No.165073	1.29	1.23
No.168198	4.53	2.97
No.171723	2.13	2.39
No.171819	1.56	1.81
No.181432	2.03	2.40
No.182368	2.69	3.00
No.183665	2.88	2.83
No.189474	2.00	1.62
No.189640	1.85	1.82
No.191294	3.25	3.54
No.194936	2.11	2.49
No.195218	1.75	2.00
No.195831	2.88	2.98
No.196462	2.06	2.58
No.197248	2.09	2.55
No.197372	1.78	1.96
No.198637	2.24	2.51
No.200737	2.13	2.68
No.206685	1.76	2.39
No.209705	1.52	1.81
No.211156	2.36	2.59
No.214617	2.21	2.78
No.251327	2.85	3.15
No.260544	3.79	4.12
No.264785	3.70	3.58
No.268508	2.58	2.87
No.268949	3.10	2.82
No.272299	2.83	2.60
No.272350	2.03	2.54
No.274971	2.38	2.51
No.466151	2.22	1.94
No.473771	1.67	2.39
No.484140	2.34	2.30
No.492284	2.05	1.88
No.550212	1.91	2.14
Adiponectin	5.48	1.94

Phosphorylation of AMPK normalized to the amount of AMPK in C2C12 myotubes and transfected with or without the indicated siRNA duplex, treated for 5 min with adiponectin or the indicated small molecule.

LETTER

Puzzling accretion onto a black hole in the ultraluminous X-ray source M 101 ULX-1

Ji-Feng Liu[1], Joel N. Bregman[2], Yu Bai[1], Stephen Justham[1] & Paul Crowther [3]

There are two proposed explanations for ultraluminous X-ray sources[1,2] (ULXs) with luminosities in excess of 10^{39} erg s^{-1}. They could be intermediate-mass black holes (more than 100–1,000 solar masses, $M_\odot$) radiating at sub-maximal (sub-Eddington) rates, as in Galactic black-hole X-ray binaries but with larger, cooler accretion disks[3–5]. Alternatively, they could be stellar-mass black holes radiating at Eddington or super-Eddington rates[2,6]. On its discovery, M 101 ULX-1[4,7] had a luminosity of 3×10^{39} erg s^{-1} and a supersoft thermal disk spectrum with an exceptionally low temperature—uncomplicated by photons energized by a corona of hot electrons—more consistent with the expected appearance of an accreting intermediate-mass black hole[3,4]. Here we report optical spectroscopic monitoring of M 101 ULX-1. We confirm the previous suggestion[8]

that the system contains a Wolf-Rayet star, and reveal that the orbital period is 8.2 days. The black hole has a minimum mass of $5M_\odot$, and more probably a mass of $20M_\odot-30M_\odot$, but we argue that it is very unlikely to be an intermediate-mass black hole. Therefore, its exceptionally soft spectra at high Eddington ratios violate the expectations for accretion onto stellar-mass black holes[9–11]. Accretion must occur from captured stellar wind, which has hitherto been thought to be so inefficient that it could not power an ultraluminous source[12,13].

Although it is desirable to obtain the primary mass of a ULX through measuring the motion of its companion (the secondary), this is only possible in the X-ray-low state (that is, at low X-ray luminosities) because the X-ray irradiated accretion disk will dominate optical emission in the X-ray-high state[14,15]. We performed a spectroscopic monitoring
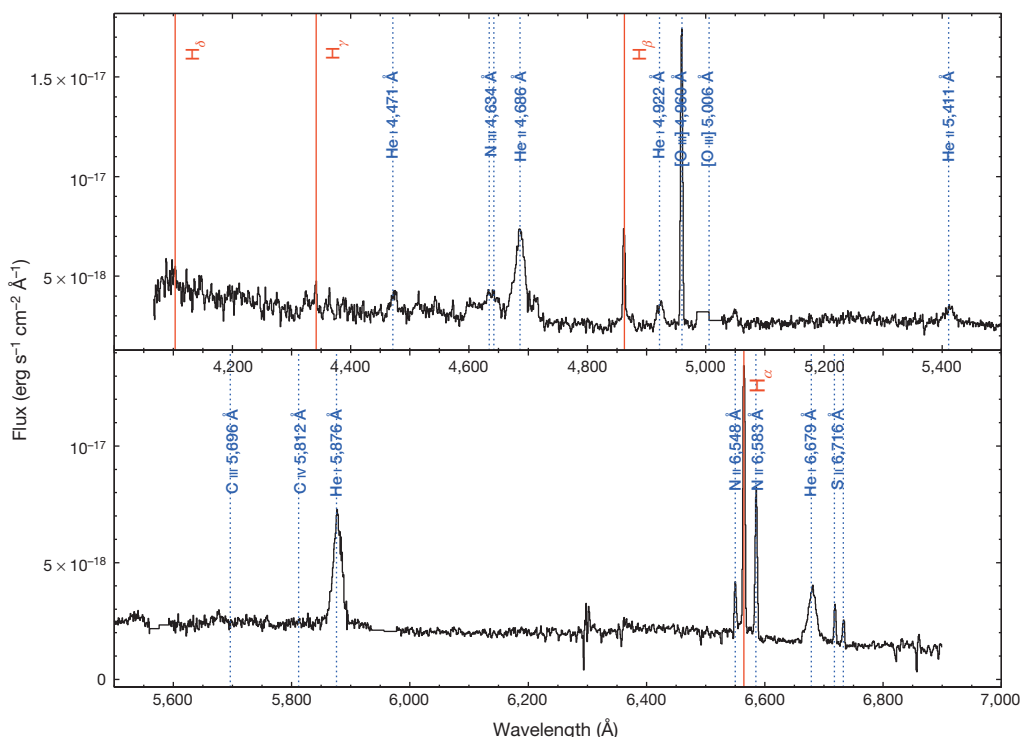
Figure 1 | The secondary of M 101 ULX-1 is confirmed to be a Wolf-Rayet star. Confirmation is based on the optical spectrum, combined from 10 Gemini/GMOS observations with a total exposure time of 16 h. The spectrum shows narrow nebular lines with a full-width at half-maximum (FWHM) of ~4 Å at the instrumental spectral resolution, including hydrogen Balmer lines and forbidden lines such as [O III] 4,960/5,006 Å (the latter is mostly in the CCD gap and only partly shown), [N II] 6,548/6,583 Å and [S II] 6,716/6,731 Å, all at a constant radial velocity over observations consistent with that of M 101. Also present are broad emission lines with FWHM up to 20 Å, including strong He II 4,686 Å, He I 5,876 Å and He I 6,679 Å lines, weaker He I 4,471 Å, He I 4,922 Å and He II 5,411 Å lines and N III 4,640 Å lines. The observed

He I 5,876/He II 5,411 Å equivalent width ratio suggests a Wolf-Rayet star of WN8 sub-type, consistent with the absence of carbon emission lines for WC stars (such as C III 5,696 Å and C IV 5,812 Å). The intensities of the helium emission lines can be best reproduced by an atmospheric model[16] of a Wolf-Rayet star with $R_* = 10.7R_\odot$, $M_* = 17.5M_\odot$, $L_* = 5.4 \times 10^5 L_\odot$, $T_* = 48$ kK, $\dot{M}_* = 2 \pm 0.5 \times 10^{-5} M_\odot$ yr^{-1} and $v_\infty = 1,300 \pm 100$ km s^{-1} (with 68.3% uncertainties for the two continuously variable parameters; details of all parameters are given in Methods), consistent with those for a WN8 star. The mass–luminosity relation[17,18] for Wolf-Rayet stars gives a more reliable mass estimate of $19M_\odot$, which we use in the main text, with an estimated formal error of $1M_\odot$.

[1]Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, 100012 Beijing, China. [2]Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, Michigan 40185, USA. [3]Department of Physics and Astronomy, University of Sheffield, Hounsfield Road, Sheffield S3 7RH, UK.

campaign for M 101 ULX-1 from February to May 2010 during its expected X-ray-low states. The optical spectrum (Fig. 1) is characterized by broad helium emission lines, including the He II 4,686 Å line. Given the absence of broad hydrogen emission lines, which are detected in some ULXs from their X-ray irradiated accretion disk at very high luminosities[14,15], the donor cannot be hydrogen rich, and thus must be a Wolf-Rayet star or a helium white dwarf. The latter can be excluded because a white dwarf is roughly a million times dimmer than the observed optical counterpart even during the low states. Indeed, the optical spectrum is unique to Wolf-Rayet stars, and the intensities of the helium emission lines can be reproduced well by an atmospheric model[16] of such a star, the mass of which is estimated to be $19M_\odot$ on the basis of the empirical mass–luminosity relation[17,18]. Given the relatively low luminosities in the X-ray-low state, the helium emission lines are expected to originate mainly from the Wolf-Rayet secondary with little contribution from the accretion disk. Such emission lines have been used to measure the black-hole mass in both IC 10 X-1 ($21M_\odot$–$35M_\odot$)[19,20] and NGC 300 X-1 ($12M_\odot$–$24M_\odot$)[21,22], systems that exhibit luminosities an order of magnitude lower than the peak luminosity of M 101 ULX-1.

Because the centroid of the He II 4,686 Å emission line varied by ± 60 km s^{-1} over the three months of our monitoring campaign, we have been able to obtain the orbital period of $P = 8.2 \pm 0.1$ days and the mass function $f(M_*, M_\bullet, i) = 0.18M_\odot \pm 0.03M_\odot$ for M 101 ULX-1 (Fig. 2). Because we already know the mass of the donor star (M_*) we are able to infer the mass of the accretor to be $M_\bullet \geq 4.6M_\odot \pm 0.3M_\odot$ (for inclination angle $i \leq 90°$), where the error is computed from the

Figure 2 | An orbital period of ~8.2 days is revealed by radial velocity measurements taken over three months for M 101 ULX-1. a, Radial velocities of the He II 4,686 Å emission line (with 68.3% uncertainties computed mainly from the dispersion of the wavelength calibration) from nine observations over three months. **b, c**, χ^2 computed for a sine fit (under the assumption of a circular orbit) to the radial velocity curve as a function of trial periods (**b**). The trial periods range from a minimum of 3 days, when the Wolf-Rayet secondary fills its Roche lobe, to a maximum of 10 days as suggested by the last five measurements. The best fit is achieved at minimal $\chi^2 \approx 1.6$ for $P = 8.2$ days and $K = 61$ km s^{-1}, for which the folded radial velocity curve is shown in **c**. The 68.3% uncertainties for the best fit are estimated to be $\Delta P = 0.1$ days and $\Delta K = 5$ km s^{-1} using $\chi^2 - \chi^2_{best} = 1$. All other trial periods (such as those at $P \approx 6.4$ days) are worse by $\Delta\chi^2 > 4$. The successful fit with a sine curve suggests that the orbital eccentricity is small. This leads to a mass function $f(M_*, M_\bullet, i) = \frac{PK^3}{2\pi G} = 0.18M_\odot \pm 0.03M_\odot$, where the error accounts for the 68.3% uncertainties in P and K.

uncertainties in the secondary mass and in the mass function. Even for the minimum mass, obtained when the system is aligned perfectly edge-on to the line of sight (for which $i = 90°$), such a compact primary can only be a black hole. Higher black-hole masses are easily obtained for lower inclination angles. For example, a stellar-mass black hole of $20M_\odot$ corresponds to $i = 19°$, and an intermediate-mass black hole (IMBH) of $1,000M_\odot$ ($300M_\odot$) corresponds to $i = 3°$ ($i = 5°$). The probability of discovering a pole-on binary with $i < 3°$ ($i = 5°$) by mere chance is lower than 0.1% (0.3%). This makes it very unlikely that this system contains an IMBH of $1,000M_\odot$ ($300M_\odot$). If the peak luminosity of M 101 ULX-1 corresponds to less than 30% of the Eddington level—which is commonly assumed to be required to produce the thermally dominated spectral state[9,23]—then the black-hole mass would exceed $50M_\odot$–$80M_\odot$. The true black-hole mass seems likely to be ~$20M_\odot$–$30M_\odot$ (see Methods for details).

The confirmation of a Wolf-Rayet star in the system, independent of the dynamical mass measurement, also suggests that M 101 ULX-1 is unlikely to be an IMBH. IMBHs cannot form directly through the collapse of massive stars, but it is suggested that they can form through mergers in dense stellar environments[24,25]. However, any IMBH formed would not be seen as a ULX unless they capture a companion as a reservoir from which to accrete matter. Such a capture is a rare event even in dense stellar environments such as globular clusters or galactic bulges, to which M 101 ULX-1 apparently does not belong, and captures that can provide high-enough accretion rates to power a ULX are even more unusual[26,27]. Given the rarity of Wolf-Rayet stars (there are about 2,000 such stars out of the 200 billion stars in a typical spiral galaxy like the Milky Way[18]), it is extremely unlikely that M 101 ULX-1 is such a revived IMBH. Alternatively a huge population of IMBHs could somehow remain undetected, both with and without companions.

M 101 ULX-1 is thus a stellar black hole, although it is a member of the class of supersoft ULXs which have been considered to be outstanding IMBH candidates[4,5]. Its combination of high luminosities and low disk temperatures (Fig. 3) strains our current understanding of accretion by stellar-mass black holes[9–11]. Studies of Galactic black-hole X-ray binaries suggest that radiation at less than roughly 30% of the Eddington luminosity is dominated by the thermal emission from a hot disk (~1 keV). A hard power-law component due to Comptonization by the disk corona becomes more and more significant when the luminosity increases to near-Eddington levels. When the luminosity increases further, to Eddington or super-Eddington levels, the Comptonized component begins to dominate the disk component, as observed for ULXs in the ultraluminous state[2,6]. For example, the ultraluminous microquasar in M31 with a stellar-mass black hole (~$10M_\odot$) and a luminosity of 10^{39} erg s^{-1} exhibited hard X-ray spectra[28]. If it were the same phenomenon, a hard X-ray spectrum would be expected for a stellar-mass black hole in M 101 ULX-1, whether it is radiating at sub-, near- or super-Eddington luminosities. The observed supersoft X-ray spectra lack hard photons above 1.5 keV, and can be described purely by cool accretion disks, uncomplicated by Comptonization, with exceptionally low temperatures of 90–180 eV (refs 4, 7). Including extra photoelectric absorption by the surrounding Wolf-Rayet wind in the spectral analysis would further lower the underlying disk temperatures and increase the luminosities[4], which would cause M 101 ULX-1 to deviate even farther from the expected hard spectra. This unambiguously demonstrates that stellar-mass black holes can have very cool accretion disks uncomplicated by the Comptonized component, contrary to standard expectations[3,9,11].

M 101 ULX-1 is the third known Wolf-Rayet/black-hole binary but is distinctly different from NGC 300 X-1 and IC 10 X-1. Whereas M 101 ULX-1 is a recurrent transient with supersoft spectra and low disk temperatures, both IC 10 X-1 and NGC 300 X-1 show constant X-ray output (despite apparent variations due to orbital modulation), hard spectra with a minor disk component, and disk temperatures above 1 keV (refs 19, 21, 29; Fig. 3). Hence the compact object in M 101 ULX-1 was considered to be an excellent IMBH candidate, whereas IC 10 X-1
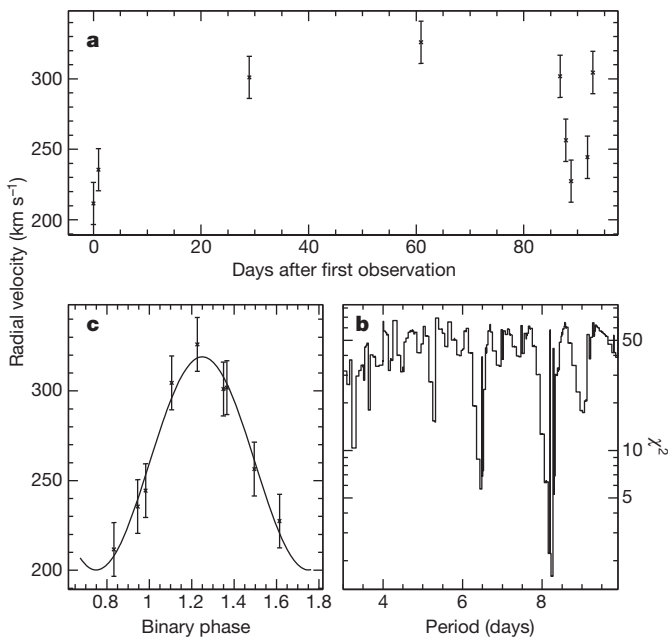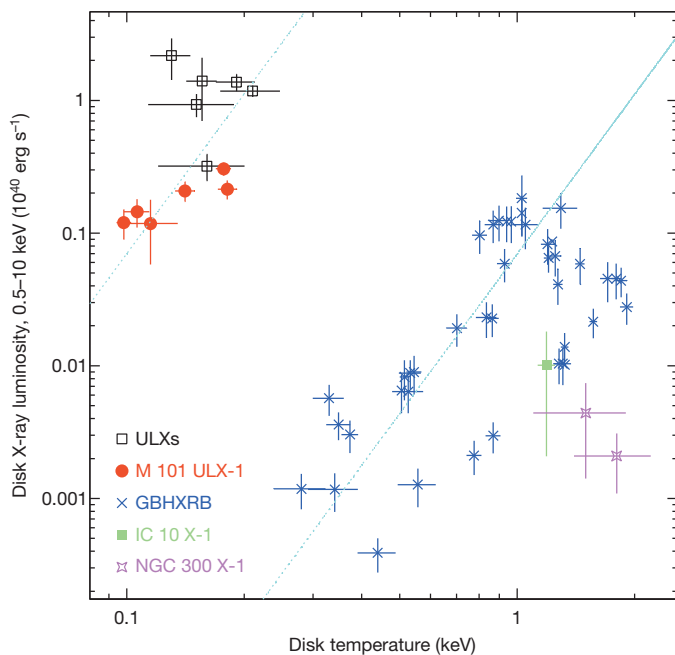
Figure 3 | The prototype ultraluminous supersoft X-ray source M 101 ULX-1 exhibits distinct spectral characteristics. M 101 ULX-1 is compared to Galactic black-hole X-ray binaries (GBHXRBs), Wolf-Rayet/black hole binaries IC 10 X-1 and NGC 300 X-1, and other ULXs on the disk X-ray luminosity (L_X) versus disk temperature (T_d) plane, all plotted with the 68.3% uncertainties from the X-ray spectral fitting. Except for M 101 ULX-1, which can be fitted with a disk blackbody model with temperatures of 90–180 eV (refs 4, 7), all other X-ray sources are complicated by the presence of a hard power-law component due to Comptonization by a corona, and can be best fitted with a disk blackbody plus power-law composite model[3,29]. Whereas GBHXRBs[3] and the other two Wolf-Rayet/black-hole binaries[29] with stellar black holes cluster in the same region, M 101 ULX-1 lies within a distinct region that has been expected to contain IMBH candidates, the same region as for some ULXs[3]. The dotted lines describe the expected disk luminosity (L_d) for different disk temperatures for a fixed disk inner radius (R_{in}) based on the relation $L_d \propto R_{in}^2 T_d^4$. The two lines are offset by four orders of magnitude in luminosity, implying a factor of 100 differences in the disk inner radii, and a factor of 100 differences in the black-hole masses if the disk radius is tied to the innermost stable orbit of the black hole. Fitting ULX spectra with alternative Comptonization models can yield high disk temperatures consistent with those of stellar-mass black holes[6]. However, the location of M 101 ULX-1 on the $L_X - T_d$ plane does not change because its spectra are not complicated by Comptonization.

and NGC 300 X-1 were expected to host stellar-mass black holes (as was later confirmed). The 8.2-day orbital period shows that M 101 ULX-1 is a wide binary, with components that would be separated by $50R_\odot$ for black-hole mass $M_\bullet = 5M_\odot$ ($75R_\odot$ for $M_\bullet = 60M_\odot$). The Roche lobe radius for the secondary is always greater than $22R_\odot$, twice as large as the Wolf-Rayet star itself. Mass transfer by Roche lobe overflow is thus impossible, and the black hole must be accreting matter by capturing the thick stellar wind. Given the geometry of the system, the disk is very large, and thus there will be a helium partial ionization zone. Such a disk is prone to instability, causing the observed X-ray transient behaviours for M 101 ULX-1. In contrast, IC 10 X-1 and NGC 300 X-1 have shorter orbital periods (34.9 h and 32.3 h respectively) and smaller separations ($\sim 20R_\odot$). Because those Wolf-Rayet stars fill their Roche lobes, the black holes accrete via Roche-lobe overflow. These systems also have much smaller and hotter accretion disks without helium partial ionization zones, which explains why IC 10 X-1 and NGC 300 X-1 do not display disk-instability outbursts (see also Methods).

Mass transfer through wind accretion usually has a very low efficiency, as in the case of many low-luminosity, high-mass X-ray binaries, and is typically not considered for populations that require high accretion

rates. However, M 101 ULX-1 demonstrates that this expectation is not always correct. In particular, transient outbursts of such wind-accreting system have generally not been included in theoretical ULX populations[12,13], but M 101 ULX-1 does attain ULX luminosities. Theorists have recently suggested that wind accretion may potentially also be significant for some progenitors of type Ia supernovae[30]. M 101 ULX-1 empirically supports this reassessment of the potential importance of wind accretion.

METHODS SUMMARY

Analysis of earlier M 101 ULX-1 observations, data reduction and analysis of the Gemini/GMOS spectroscopic observations, determination of the Wolf-Rayet sub-class and its physical parameters, the search for orbital periodicity, and determination of the properties of the Wolf-Rayet/black-hole binary are described in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Fabbiano, G. The hunt for intermediate-mass black holes. *Science* **307,** 533–534 (2005).
2. Gladstone, J. The sub-classes of ultraluminous X-ray sources. Preprint at http://arXiv.org/abs/1306.6886 (2013).
3. Miller, J. *et al.* A comparison of intermediate mass black hole candidate ultraluminous X-ray sources and stellar mass black holes. *Astrophys. J.* **614,** L117–L120 (2004).
4. Kong, A. K. H., Di Stefano, R. & Yuan, F. Evidence of an intermediate-mass black hole: Chandra and XMM-Newton observations of the ultraluminous supersoft X-ray source in M101 during its 2004 outburst. *Astrophys. J.* **617,** L49–L52 (2004).
5. Liu, J. F. & Di Stefano, R. An ultraluminous supersoft X-ray source in M81: an intermediate-mass black hole? *Astrophys. J.* **674,** L73–L77 (2008).
6. Gladstone, J. *et al.* The ultraluminous state. *Mon. Not. R. Astron. Soc.* **397,** 1836–1851 (2009).
7. Mukai, K., Still, M., Corbet, R., Kuntz, K. & Barnard, R. The X-ray properties of M101 ULX-1 = CXOKM101 J140332.74+542102. *Astrophys. J.* **634,** 1085–1092 (2005).
8. Liu, J. F. Multi-epoch multi-wavelength study of an ultraluminous X-ray source in M101: the nature of the secondary. *Astrophys. J.* **704,** 1628–1639 (2009).
9. McClintock, J. & Remillard, R. in *Compact Stellar X-ray Sources* (eds Lewin, W. & van der Klis, M.) 157–213 (Cambridge Astrophysics Series No. 39, Cambridge Univ. Press, 2006).
10. Esin, A. A., McClintock, J. E. & Narayan, R. Advection-dominated accretion and the spectral states of black hole X-ray binaries: application to nova Muscae 1991. *Astrophys. J.* **489,** 865–889 (1997).
11. Remillard, R. A. & McClintock, J. E. X-ray properties of black-hole binaries. *Annu. Rev. Astron. Astrophys.* **44,** 49–92 (2006).
12. Rappaport, S., Podsiadlowski, Ph. & Pfahl, E. Stellar-mass black hole binaries as ultraluminous X-ray sources. *Mon. Not. R. Astron. Soc.* **356,** 401–414 (2005).
13. Linden, T. *et al.* The effect of starburst metallicity on bright X-ray binary formation pathways. *Astrophys. J.* **725,** 1984–1994 (2010).
14. Roberts, T. P. *et al.* (No) dynamical constraints on the mass of the black hole in two ULXs. *Astron. Nachr.* **332,** 398–401 (2011).
15. Cseh, D., Gris'e, F., Corbel, S. & Kaaret, P. Broad components in optical emission lines from the ultraluminous X-ray source NGC 5408 X-1. *Astrophys. J.* **728,** L5–L9 (2011).
16. Hillier, D. J. & Miller, D. L. The treatment of non-LTE line blanketing in spherically expanding outflows. *Astrophys. J.* **496,** 407–427 (1998).
17. Schaerer, D. & Maeder, A. Basic relations between physical parameters of Wolf-Rayet stars. *Astron. Astrophys.* **263,** 129–136 (1992).
18. Crowther, P. A. Physical properties of Wolf-Rayet stars. *Annu. Rev. Astron. Astrophys.* **45,** 177–219 (2007).
19. Prestwich, A. H. *et al.* The orbital period of the Wolf-Rayet binary IC 10 X-1: dynamic evidence that the compact object is a black hole. *Astrophys. J.* **669,** L21–L24 (2007).
20. Silverman, J. M. & Filippenko, A. V. On IC 10 X-1, the most massive known stellar-mass black hole. *Astrophys. J.* **678,** L17–L20 (2008).
21. Carpano, S. *et al.* A 33 hour period for the Wolf-Rayet/black hole X-ray binary candidate NGC 300 X-1. *Astron. Astrophys.* **466,** L17–L20 (2007).
22. Crowther, P. A. *et al.* NGC 300 X-1 is a Wolf-Rayet/black hole binary. *Mon. Not. R. Astron. Soc.* **403,** L41–L45 (2010).
23. Steiner, J. F., McClintock, J. E., Remillard, R. A., Narayan, R. & Gou, L. J. Measuring black hole spin via the X-ray continuum-fitting method: beyond the thermal dominant state. *Astrophys. J.* **701,** L83–L86 (2009).
24. Miller, M. C. & Hamilton, D. P. Production of intermediate-mass black holes in globular clusters. *Mon. Not. R. Astron. Soc.* **330,** 232–240 (2002).
25. Portegies Zwart, S. F., Baumgardt, H., Hut, P., Makino, J. & McMillan, S. L. W. Formation of massive black holes through runaway collisions in dense young star clusters. *Nature* **428,** 724–726 (2004).
26. Blecha, L. *et al.* Close binary interactions of intermediate-mass black holes: possible ultra-luminous X-ray sources? *Astrophys. J.* **642,** 427–437 (2006).

27. Madhusudhan, N. *et al.* Models of ultraluminous X-ray sources with intermediate-mass black holes. *Astrophys. J.* **640,** 918–922 (2006).
28. Middleton, M. J. *et al.* Bright radio emission from an ultraluminous stellar-mass microquasar in M31. *Nature* **493,** 187–190 (2013).
29. Barnard, R., Clark, J. S. & Kolb, U. C. NGC 300 X-1 and IC 10 X-1: a new breed of black hole binary? *Astron. Astrophys.* **488,** 697–703 (2008).
30. Mohamed, S. & Podsiadlowski, Ph. Mass transfer in Mira-type binaries. *Balt. Astron.* **21,** 88–96 (2011).

METHODS

Analysis of earlier M 101 ULX-1 observations. M 101 is a nearby face-on grand design spiral galaxy, a frequent target of various observations. These include the optical monitoring observations in search of Cepheids with the Hubble Space Telescope, yielding a distance of 6.855 Mpc (ref. 31). M 101 ULX-1 (CXO J140332.3+542103) is located near a spiral arm (Extended Data Fig. 1), and identified with a unique optical counterpart of $V = 23.5$ mag (ref. 32). At this location, the metallicity is 0.4 times solar according to the M 101 gas-phase oxygen abundance gradient[33].

This ULX has been observed intensively by X-ray missions including ROSAT, XMM and Chandra since early 1990s, which exhibited spectral state transitions between the low-hard state and the high-soft state reminiscent of Galactic black-hole X-ray binaries. This ULX was once the brightest X-ray point source in M 101 with a Chandra/ACIS count rate of 0.10 counts s^{-1} (ref. 34), observed during the 2000 March observation (ObsID 934). The Chandra/XMM-Newton spectra during its outbursts[4,35] were very soft and can be generally fitted with an absorbed blackbody model with neutral hydrogen column density $n_{\rm H} = (1-4) \times 10^{21}$ cm^{-2} and temperatures of 50–100 eV, and the peak 0.3–7 keV luminosity reached 3×10^{40} erg s^{-1}, with a bolometric luminosity of about 10^{41} erg s^{-1}, suggesting an IMBH of a few thousand solar masses. It was argued that it is unphysical to adopt a high neutral hydrogen column density of $\gtrsim 10^{21}$ cm^{-2}, and fitting the spectra as blackbody plus a diskline component centred at 0.5 keV with $n_{\rm H}$ fixed at the Galactic value of 4×10^{20} cm^{-2} yielded the maximum outburst bolometric luminosity of 3×10^{39} erg s^{-1}, consistent with the Eddington luminosity of a black hole of $20M_\odot$–$40M_\odot$ (ref. 7).

Even at the lowered luminosities of 3×10^{39} erg s^{-1}, the combination of the disk luminosities and disk temperatures makes M 101 ULX-1 an outstanding IMBH candidate. It is believed that the accretion disks for IMBHs should have larger inner radii and consequently lower disk temperatures[3–5], occupying the upper left portion in the $T_{\rm d}$–$L_{\rm X}$ plane as shown in Fig. 3. The position of M 101 ULX-1 on this plane suggests that it is distinctly different from the Galactic black-hole X-ray binaries in the lower right portion, but belongs to the league of IMBH candidates along with some extreme ULXs above 10^{40} erg s^{-1}. The practice of placing these ULXs on this plane was questioned because decomposing ULX spectra into disk blackbody plus power-law models is unphysical given the dominance of the hard power-law component. However, in the case of M 101 ULX-1 the spectra are supersoft without any hard power-law component, so its location on the plane should reflect the accretion disk uncomplicated by Comptonization. For comparison, we also put on this plane the other two known Wolf-Rayet/black-hole binaries[29] IC 10 X-1 and NGC 300 X-1, which apparently belong to the league of stellar-mass black holes, and dynamical mass measurements have yielded mass estimates of $20M_\odot$–$30M_\odot$.

Combined analysis of 26 HST observations and 33 X-ray observations over 16 years[8] revealed two optical outbursts in addition to 5 X-ray outbursts. Although there is no 'exact' period for the recurring outbursts, the outbursts occur once roughly every six months. Such outbursts last 10–30 days, suggesting an outburst duty cycle of 10–15%. Outside outbursts, ULX-1 stays in a low-hard state with an X-ray luminosity of 2×10^{37} erg s^{-1} (refs 4, 7, 8, 35). Such behaviour is reminiscent of those of soft X-ray transients in low-mass X-ray binaries, albeit with higher luminosities and lower disk temperatures, but is different from the recently discovered high-mass fast transients owing to clumping winds at much lower X-ray luminosities ($\sim 10^{34}$ erg s^{-1}). Detailed studies of the optical spectral energy distribution, after removal of optical emission from the X-ray irradiated accretion disk in the outbursts, suggest that the secondary is a Wolf-Rayet star of initially $40M_\odot$–$60M_\odot$, currently $18M_\odot$–$20M_\odot$, $9R_\odot$–$12R_\odot$ and about 5×10^4 K (ref. 8). This claim of a Wolf-Rayet companion is supported by the presence of the He II 4,686 Å emission line in the Gemini/GMOS-N spectrum taken in 2005[36].

Gemini/GMOS data reduction. M 101 ULX-1 was monitored spectroscopically from February to May in 2010 during its expected low states under the Gemini/GMOS-N program GN-2010A-Q49 (PI: J.-F.L.). Extended Data Table 1 lists the observations taken in ten nights distributed from February to May, with a total exposure of 15.6 h. All exposures were taken with the 0.75″ slit and the B600 grating tuned for a wavelength coverage from 4,000–6,900 Å; such a slit/grating combination will yield a spectral resolution of about 4.5 Å. We followed standard procedures to reduce the observations and extract 1D spectra using the gmos package in IRAF. All consecutive sub-exposures during one night were combined into one spectrum to increase the signal-to-noise ratio, and we obtained ten spectra with exposure times ranging from 3,200 s to 9,600 s (Extended Data Table 1).

For each spectrum, the wavelength solution was obtained using the copper–argon arc lamp spectra taken with the same slit/grating setting right before and after the science exposures during the same night or occasionally the night after. We verified the wavelength solution by comparing thus-obtained wavelengths to the intrinsic wavelengths for a dozen of strong night sky emission lines identified in the spectra before sky subtraction, and revealed wavelength differences with a

dispersion of about 0.25 Å or ~ 15 km s^{-1}. The extracted spectra were converted to flux spectra using the standard star HZ44 taken during the night of February 15, and we scaled the spectra to have specific flux $f_\lambda = 1.5 \times 10^{-18}$ erg s^{-1} cm^{-2} Å^{-1} at 5,500 Å corresponding to F555W = 23.5 mag based on previous HST/WFPC2 observations[8].

Figure 1 shows the flux-calibrated sky-subtracted spectrum combined from the ten spectra. The combined spectrum is free of absorption lines but abundant in emission lines as identified and listed in Extended Data Table 2. For each emission line, we fit a Gaussian profile to derive its line width and compute its line flux and luminosity. Two categories of lines are present in the spectrum. The first category is the broad helium emission lines with FWHM of up to 20 Å, five times broader than the instrumental spectral resolution, and includes strong He II 4,686 Å, He I 5,876 Å, He I 6,679 Å, and weaker He I 4,471 Å, He I 4,922 Å, and He II 5,411 Å lines. The broad N III 4,634 Å emission line is also present. The second category is the narrow emission lines with line widths consistent with the instrumental spectral resolution, and includes the Balmer lines and forbidden lines such as [O III] 4,960/5,006 Å (the latter is mostly in the CCD gap and not listed), [N II] 6,548/6,583 Å, and [S II] 6,716/6,731 Å.

The emission line properties are derived from the Gaussian line profile fitting. The average line properties including FWHM, equivalent width, and line luminosities are measured from the combined spectrum (Extended Data Table 2). The shifts of the line centres were also measured for individual spectra, with the barycentric correction computed using the rvsao package in IRAF as listed in Extended Data Table 1 for each spectrum. It was found that the line shifts, after barycentric correction, are consistent with being constant for narrow emission lines over all observations at 230 ± 15 km s^{-1}, consistent with the radial velocity of 241 ± 2 km s^{-1} for the face-on M 101. However, the broad helium emission lines, as measured with the strongest He II 4,686 Å line, shifted from observation to observation between 210 km s^{-1} and 330 km s^{-1} as listed in Extended Data Table 1, with an average of 270 km s^{-1} that is significantly different from that for nebular lines.

The properties of the nebular lines help to determine the environmental metallicity and the neutral hydrogen column density. The line intensity ratio between [N II] 6,583 Å and H$_\alpha$, $N2 \equiv$ [N II]λ6,583$/$H$_\alpha$, can be used as an abundance indicator[37] with $12 + \log(\rm O/H) = 8.90 + 0.57 \times N2$, albeit with a large dispersion in $\log(\rm O/H)$ of ± 0.41. Given the equivalent width of these two lines (Extended Data Table 2), we find $12 + \log(\rm O/H) = 8.70$, close to solar metallicity (8.66). This is higher than but marginally consistent with the value of 0.4 times solar according to the M 101 gas-phase oxygen abundance gradient[33] given the location of ULX-1. The observed Balmer line flux ratios can be used to infer the dust extinction between the nebula and the observer. In the nebular emission around ULX-1, the intrinsic ratio H$_\alpha$/H$_\beta$ is 2.74 in case B for a thermal temperature of $T = 20,000$ K (ref. 38). Assuming $E(B-V) = 0.1$ mag, then $A_{6564} = 0.250$ mag, $A_{4863} = 0.360$ mag, $\Delta A = 0.11$ mag, ΔH$_\alpha/\Delta$H$_\beta = 1.1$, and reddened H$_\alpha$/H$_\beta \approx 3$. The observed H$_\alpha$/H$_\beta$ is 2.85, suggesting that the extinction is low, and using the Galactic value is reasonable.

Determining the Wolf-Rayet subclass of binary ULX-1. The broad helium emission lines in the newly obtained Gemini/GMOS spectrum are typical of an extremely hot, hydrogen-depleted Wolf-Rayet star. Accretion disks around a compact object can also give rise to broad helium emission lines, but a broad Balmer line is expected to be present and much stronger than the helium lines. Indeed, broad H$_\beta$ emission lines are present in two ULXs with optical spectra (4,000–5,400 Å), NGC1313 X-2[14] and NGC 5408 X-1[15], and are stronger than the He II 4,686 Å emission line. In the ULX-1 spectrum (Fig. 1), although the Balmer emission lines are present, they are narrow emission lines like forbidden lines, and should come from the surrounding nebulae, as evidenced by their nearly constant line shifts from observation to observation, in distinct contrast to helium lines with line shift differences of ± 60 km s^{-1}.

The sub-type of this Wolf-Rayet star can be determined from the presence or absence of line species in the spectrum[18]. There are two main types of Wolf-Rayet stars, WN stars with $R \approx 5R_\odot$–$12R_\odot$ revealing H-burning products, and subsequently more compact WC stars with R $\approx 2R_\odot$–$3R_\odot$ revealing He-burning products. Spectra of WC stars are dominated by carbon lines (such as C III 4,650 Å, C III 5,696 Å and C IV 5,812 Å) that are stronger than helium lines, but none of the carbon lines are present in the ULX-1 spectrum. WN stars from WN4 to WN8 show[39] increasing absolute magnitudes M_V from -3.5 mag to -6 mag, increasing mass loss rates from $10^{-5}M_\odot$ yr^{-1} to $10^{-4}M_\odot$ yr^{-1}, decreasing effective temperatures from 80 kK to 45 kK, and hence an increasing fraction of He I atoms relative to He II ions. Comparing the observed spectrum to the spectral atlas of WN stars[18,40], we estimate a late-type WN8 star. A WN8 subtype is also inferred based on the He I 5,876 Å /He I 5,411 Å equivalent width ratio[41]. Such a subtype is roughly consistent with its absolute magnitude of $M_V = -5.9$ mag (after extinction correction using Galactic $E(B-V) = 0.1$ mag and $R_V = 3.1$), and the effective temperature of about 50 kK derived from its broad-band spectral energy distribution[8].

Determining physical parameters for the Wolf-Rayet star. As for the case of NGC 300 X-1[22], we have calculated synthetic models using the line-blanketed,

non-local thermodynamic equilibrium model atmosphere code[16]. To select the best physical parameters of the Wolf-Rayet star, we compare the model equivalent width EW with observed values for the six helium emission lines and minimize the quantity $\Delta^2 = \Sigma_i(EW - EW_i)^2$. In all model calculations, elemental abundances are set to 40% of the solar value for the metallicity of $0.4Z_\odot$ at the location of ULX-1. We vary the stellar radius R_* between $4R_\odot$ and $20R_\odot$, stellar mass M_* between $5M_\odot$ and $35M_\odot$, stellar luminosity L_* in the range $(5-100) \times 10^4 L_\odot$, the outer radius for the line-forming region R_{MAX} up to $40R_\odot$, the terminal velocity v_∞ between 400 and 2,000 km s^{-1}, and the stellar wind mass loss rate $\dot{M}_*$ in the range $(5-100) \times 10^{-6} M_\odot$ yr^{-1}.

We have run ~5,000 models with the combinations of stellar mass, radius and luminosity determined by the stellar evolution tracks[42] of $Z = 0.4Z_\odot$ for all possible WN stars, and another ~5,000 models with 'fake' stars whose mass, radius and luminosity are completely independent of each other. After a total of ~10,000 model evaluations, a best-fitting model is found with $R_* = 10.7R_\odot$, $M_* = 17.5M_\odot$, $L_* = 5.4 \times 10^5 L_\odot$, $v_\infty = 1,300$ km s^{-1}, $R_{MAX} = 22R_\odot$ and $\dot{M}_* = 2.0 \times 10^{-5} M_\odot$ yr^{-1}. The model reproduces the helium emission lines extremely well (Extended Data Table 2), with an average difference of $|\Delta| = 0.6$ Å. In comparison, the majority of models and all models with 'fake' parameters are much worse-fitting with $\Delta^2 \gg 10$ (Extended Data Fig. 2). Based on the Δ^2 distribution, our model evaluations picked up the stellar parameters effectively, and we estimate, with equivalently $\Delta\chi^2 = 1$, the errors to be $\dot{M}_* = (2 \pm 0.5) \times 10^{-5} M_\odot$ yr^{-1}, and $v_\infty = 1,300 \pm 100$ km s^{-1}. Note that, if we adopt a solar metallicity, as allowed by the abundance indicator $N2 \equiv [N \text{ II}]\lambda6,583/H_\alpha$, the best model will change to $R_* = 11.1R_\odot$, $M_* = 17.5M_\odot$, $L_* = 4.9 \times 10^5 L_\odot$, $v_\infty = 1,700$ km s^{-1} and $\dot{M}_* = 2.4 \times 10^{-5} M_\odot$ yr^{-1}. This is consistent with the $0.4Z_\odot$ results within the errors except for a significantly higher terminal velocity.

The stellar parameters of this best model belong to a 'real' WN star from the stellar evolution tracks, with an effective temperature of 48 kK, an initial mass of $42M_\odot$, an age of about 5 Myr, and a remaining lifetime of about 0.3 Myr before it loses another ~$6M_\odot$ and collapses into a black hole of ~$12M_\odot$. This model is actually one of the best models derived from studies of the optical spectral energy distribution[8]. Comparing to the physical properties of Wolf-Rayet stars in the Milky Way[18], we find that T_*, L_*, $\dot{M}_*$ and v_∞ are consistent with those for a WN7/WN8 star. The absolute magnitude M_V for ULX-1 ($M_V = -5.9$ mag after extinction correction) is brighter by 0.5 mag, fully within the spread of absolute magnitudes for WN subtypes.

The mass of the Wolf-Rayet star can be more reliably estimated with the empirical mass–luminosity relation[17,18] as done for NGC 300 X-1[22]. In our case, $L_* = 5.4 \times 10^5 L_\odot$, and this corresponds to a Wolf-Rayet mass of $19M_\odot$, quite consistent with the mass for the best model. The luminosity derived for solar metallicity will correspond to a Wolf-Rayet mass of $18M_\odot$. Hereafter we will use $19M_\odot$ for the Wolf-Rayet mass, with an estimated formal error of $1M_\odot$ to roughly reflect the difference between the model value and the empirical value. Given the stellar mass and radius of $10.7R_\odot$, we can obtain the orbital period[43] from its mean density ρ as $P = \sqrt{\rho/100}$ h ≈ 72 h if the Wolf-Rayet star is filling its Roche lobe. The true orbital period will be longer than 72 h if the Wolf-Rayet star is only filling part of its Roche lobe.

Searching for orbital periodicity. The radial velocity changes between 210 km s^{-1} and 330 km s^{-1} as measured by the He II 4,686 Å emission line should reflect the orbital motion of the Wolf-Rayet star. Although a broad He II 4,686 Å emission line can be produced from the X-ray heated accretion disk in some ULXs with rather high X-ray luminosities (for example, in NGC 1313 X-2 with ~10^{40} erg s^{-1}; ref. 14), this should not be the case for M 101 ULX-1 because its X-ray luminosities during the Gemini/GMOS observations were three orders of magnitude lower, and the disk heating effects are insignificant even in its outburst, based on the optical studies[8]. In addition, the line ratios for the heated accretion disk are different from the line ratios for the Wolf-Rayet star because the emission line forming regions and temperature structures are quite different, yet the observed line ratios can be well reproduced by the Wolf-Rayet star.

In order to search for the orbital periodicity, we assume a circular orbit and fit a sine curve $v_r = v_0 + K \sin[2\pi(t - t_1)/P + \varphi]$ to nine barycentre-corrected radial velocities; the radial velocity for March 17 was dropped from the analysis because the spectrum had a very low signal-to-noise ratio. The four parameters are the radial velocity of the binary mass centre v_0, the radial velocity semi-amplitude K, the orbital period P, and phase φ at the first observation. The search is carried out by minimizing χ^2 defined as $\chi^2 = \sum_{i=1}^{10} [v_r(t_i) - v_{r,i}]^2 / \sigma_{vr,i}^2$. The radial velocity errors $\sigma_{vr,i}$ are taken as the wavelength calibration error of 0.25 Å, or 15 km s^{-1}. The five radial velocity measurements from 13 May to 19 May suggest a period no longer than 10 days (Fig. 2). The 'amoeba' technique is used for χ^2 minimization, using initial guesses taken from the parameter grids with P from 3 to 10 days in steps of 0.01 days, K from 20 to 150 km s^{-1} in steps of 5 km s^{-1}, and φ from 0° to 360° in steps of 10°.

The best solution is found at the minimum $\chi^2 = 1.6$, for which the best period $P = 8.24 \pm 0.1$ days and the best radial velocity semi-amplitude $K = 61 \pm 5$ km s^{-1}, with the 68.3% error determined with $\Delta\chi^2 = 1$. The fact that the radial velocity curve can be fitted with a sine curve suggests that the orbital eccentricity is small.

Given P and K, the mass function for M 101 ULX-1 can be computed as $f(M_*, M_\bullet, i) = \dfrac{PK^3}{2\pi G} = \dfrac{M_\bullet^3}{(M_* + M_\bullet)^2} \sin^3 i = 0.178 M_\odot$. This sets an absolute lower limit for the mass of the primary. In the case of ULX-1, more information can be extracted because we already know $M_* = 19M_\odot$. Given the equation $\dfrac{M_\bullet^3}{(M_* + M_\bullet)^2} \sin^3 i = 0.178 M_\odot$, the primary mass will increase monotonically when the inclination angle decreases, that is, changing from edge-on ($i = 90°$) towards face-on ($i = 0°$). Thus the minimum mass for the primary can be obtained when $i = 90°$, which is $M_\bullet = 4.6M_\odot$ after solving the equation $\dfrac{M_\bullet^3}{(M_* + M_\bullet)^2} \sin^3 i = 0.178 M_\odot$. The minimum mass will be $M_\bullet = 4.4M_\odot$ if we use $M_* = 17.5M_\odot$. Such a compact primary can only be a black hole. This is thus the dynamical evidence for a black hole in a ULX.

Determining the properties of the Wolf-Rayet/black-hole binary. This section duplicates some text from the main article, but with additional technical details.

M 101 ULX-1 is thus a Wolf-Rayet/black-hole binary, only the third discovered so far after IC 10 X-1 and NGC300 X-1. The binary separation can be computed with Kepler's law $a^3 = \dfrac{G(M_* + M_\bullet)}{4\pi^2} P^2$, which increases monotonically for increasing black-hole mass, starting from $a = 50R_\odot$ for $M_\bullet = 4.6M_\odot$ to $a = 75R_\odot$ for $M_\bullet = 60M_\odot$ (Extended Data Fig. 3). The Roche lobe size for the secondary can be computed with $R_{cr} = af(q) = a0.49q^{2/3}/[0.6q^{2/3} + \ln(1 + q^{1/3})]$ with $q = M_*/M_\bullet$, and the Roche lobe size for the black hole can be computed with the same formula but with different $q = M_\bullet/M_*$. As shown in Extended Data Fig. 3, the Roche lobe size for the black hole increases with the increasing black-hole mass, but the Roche lobe size for the secondary does not change much, from $R_{cr,*} = 25R_\odot$ for $M_\bullet = 4.6M_\odot$ to $R_{cr,*} = 23R_\odot$ for $M_\bullet = 10M_\odot$, and to $R_{cr,*} = 22R_\odot$ for $M_\bullet = 20M_\odot$.

Regardless of the black-hole mass, the secondary is filling only half of its Roche lobe by radius, and the black hole must be accreting from the Wolf-Rayet star winds. Because the black hole is at least $50R_\odot$ away from the Wolf-Rayet star, the stellar wind must have reached close to its terminal velocity. The capture radius for the wind accretion can be computed as $r_{acc} = \dfrac{2GM_\bullet}{v_\infty^2}$ and the accretion rate can be computed as $\dot{M}_\bullet = \dfrac{\pi r_{acc}^2}{4\pi a^2} \dot{M}_*$. Given that the average luminosity for M 101 ULX-1 is about 3×10^{38} erg s^{-1}, the required accretion rate is $\dot{M}_\bullet = L / \eta c^2 \approx \dfrac{L_{38}}{\eta} 2 \times 10^{-9} M_\odot$ yr$^{-1} = \dfrac{1}{\eta} 6 \times 10^{-9} M_\odot$ yr^{-1}. To capture this much stellar wind matter, as shown in Extended Data Fig. 4, the black-hole mass must be greater than $46M_\odot$ for $\eta = 0.06$ in the case of a non-spinning Schwarzschild black hole, and greater than $13M_\odot$ for $\eta = 0.42$ in the case of a maximally spinning Kerr black hole. If we use the velocity law $v(r) = v_\infty(1 - R_*/r)\beta$ with $\beta = 1$ for the inner wind[16], then the black-hole mass must be greater than $28M_\odot$ for $\eta = 0.06$ in the case of a non-spinning Schwarzschild black hole, and greater than $8M_\odot$ for $\eta = 0.42$ in the case of a maximally spinning Kerr black hole. If we adopt a typical η value of 0.1, the required accretion rate corresponds to $M_\bullet > 24M_\odot$ (and $i < 17°$) for a wind velocity of $v \approx 1,100$ km s^{-1}, and corresponds to $M_\bullet > 32M_\odot$ (and $i < 14°$) for the terminal velocity. The accretion rate argument thus requires a black hole of $> 8M_\odot - 46M_\odot$, likely to be a black hole of $20M_\odot - 30M_\odot$ similar to IC 10 X-1 and NGC 300 X-1.

The recurring X-ray/optical outbursts dictate the presence of an accretion disk prone to instability, and the disk formation under stellar wind accretion places stringent constraints on the binary system. To explore why the number of Galactic X-ray stars is so small, it has been shown[44] that in the case of accretion of stellar wind matter in a detached binary system the specific angular momentum of the matter captured by the compact object is typically small. Therefore, usually no accretion disk is formed around the compact object. Consequently, very special conditions are required for a black hole in a detached binary system to be a strong X-ray source. A disk may form if the specific angular momentum of accreting matter, $Q_{acc} = \dfrac{1}{4} \dfrac{2\pi}{P} r_{acc}^2$, exceeds the specific angular momentum of the particle at the innermost stable circular orbit, $Q_{ISCO} = \sqrt{3}r_g c = \sqrt{3} \dfrac{2GM_\bullet}{c^2} c$. This is usually expressed as $P < 4.8 \dfrac{M_\bullet/M_\odot}{v_{1000}^4} \delta^2$ h, where $\delta \approx 1$ is a dimensionless parameter[19,45]. Given $P = 8.24 \pm 0.1$ days and $v_\infty = 1,300 \pm 100$ km s^{-1} for M 101 ULX-1, the black-hole mass is required to be $M_\bullet > 80M_\odot$, corresponding to $i = 9°$ (that is, nearly face-on). If the wind velocity from the velocity model of the inner wind[16] is adopted, then the black-hole mass is required to be $M_\bullet > 48M_\odot$, corresponding to $i = 11°$.

To investigate the possible presence of a partial ionization zone, we need to compute the temperature structure $T_d(r)$ for the accretion disk, especially for the outer disk. Following the procedures designed for an X-ray irradiated black-hole binary model for ULXs[46], we compute the disk temperature structure for a standard accretion disk with the α prescription[47] plus X-ray irradiation[43]. As shown in Extended Data Fig. 5, regardless of the black-hole mass for M 101 ULX-1, its outer disk temperature is as low as 4,000 K in the low-hard state owing to its large separation and large disk, and the helium partial ionization zone at about 15,000 K is bound to exist unless the black-hole mass is lower than $5.5M_\odot$. In comparison, the disk temperature for NGC 300 X-1, with an orbital period of 32.8 h and its WN5 star ($M_* = 26M_\odot$, $R_* = 7.2R_\odot$) filling its Roche lobe[22], never drops below 20,000 K owing to its small separation and small disk, and there is no helium partial ionization zone in the disk. This explains naturally why NGC 300 X-1 and similarly IC 10 X-1 exhibit steady X-ray radiation despite the apparent variations due to orbital modulation under the edge-on viewing geometry.

The existence of an accretion disk in M 101 ULX-1 is also supported by the observed spectral state changes, which resemble those for Galactic black-hole binaries[9,11] that are believed to reflect changes in the properties of their accretion disks[10]. During its outbursts, M 101 ULX-1 exhibits an X-ray spectrum[4,7] that can be classified as a thermal dominant state (albeit with exceptionally low disk temperatures), a well-defined spectral state that corresponds to a standard thin accretion disk at about 10% of its Eddington luminosity. Quantitative studies[23] show that when the luminosity exceeds 30% of the Eddington limit, the emission changes such that the X-ray spectrum includes a steep power-law with a significant hard component above 2 keV. The presence of such a hard component is not seen in the X-ray spectra of M 101 ULX-1. Given its bolometric luminosity of 3×10^{39} erg s^{-1} in the thermal dominant state at less than 30% of its Eddington limit, we infer that the black-hole mass is above $80M_\odot$. If this is true, the inferred black-hole mass of M 101 ULX-1 may challenge the expectations of current black-hole formation theories. The most massive black holes that can be produced for solar metallicity are about $15M_\odot$, and about $20M_\odot$ ($25M_\odot$, $30M_\odot$) for ×0.6 (×0.4, ×0.3) solar metallicity owing to reduced stellar winds and hence reduced mass loss in the final stages before stellar collapse[48].

31. Freedman, W. *et al.* Final results from the Hubble Space Telescope key project to measure the Hubble Constant. *Astrophys. J.* **553,** 47–72 (2001).

32. Kong, A. K. H., Rupen, M. P., Sjouwerman, L. O. & Di Stefano, R. in *Proc. Papers 22nd Texas Symp. Relativistic Astrophys. Stanford* (eds Chen, P., Bloom, E., Madejski, G. & Patrosian, V.) 606–611 (Stanford Univ. Press, 2005).

33. Bresolin, F. The oxygen abundance in the inner H II regions of M101: implications for the calibration of strong-line metallicity indicators. *Astrophys. J.* **656,** 186–197 (2007).

34. Liu, J. F. Chandra ACIS survey of X-ray point sources in 383 nearby galaxies. I. The source catalog. *Astrophys. J.* **192** (suppl.), 10–64 (2011).

35. Kong, A. K. H. & Di Stefano, R. An unusual spectral state of an ultraluminous very soft X-ray source during outburst. *Astrophys. J.* **632,** L107–L110 (2005).

36. Kuntz, K. D. *et al.* The optical counterpart of M101 ULX-1. *Astrophys. J.* **620,** L31–L34 (2005).

37. Pettini, M. & Pagel, B. E. J. [OIII]/[NII] as an abundance indicator at high redshift. *Mon. Not. R. Astron. Soc.* **348,** L59–L63 (2004).

38. Osterbrock, D. *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (University Science Books, 1989).

39. Hamann, W. R., Koesterke, L. & Wessolowski, U. Spectra analysis of the Galactic Wolf-Rayet stars — a comprehensive study of the WN class. *Astron. Astrophys.* **274,** 397–414 (1993).

40. Crowther, P. A. & Hadfield, L. J. Reduced Wolf-Rayet line luminosities at low metallicity. *Astron. Astrophys.* **449,** 711–722 (2006).

41. Smith, L. F., Shara, M. M. & Moffat, A. F. J. A three-dimensional classification for WN stars. *Mon. Not. R. Astron. Soc.* **281,** 163–191 (1996).

42. Girardi, L. *et al.* Theoretical isochrones in several photometric systems. I. Johnson-Cousins-Glass, HST/WFPC2, HST/NICMOS, Washington, and ESO imaging survey filter sets. *Astron. Astrophys.* **391,** 195–212 (2002).

43. Frank, J., King, A. & Raine, D. *Accretion Power in Astrophysics* (Cambridge Univ. Press, 2002).

44. Illarionov, A. F. & Sunyaev, R. A. Why the number of Galactic X-ray stars is so small? *Astron. Astrophys.* **39,** 185–195 (1975).

45. Ergma, E. & Yungelson, L. R. CYG X-3: can the compact object be a black hole? *Astron. Astrophys.* **333,** 151–158 (1998).

46. Liu, J. F., Orosz, J. & Bregman, J. N. Dynamical mass constraints on the ultraluminous X-ray source NGC 1313 X-2. *Astrophys. J.* **745,** 89–110 (2012).

47. Shakura, N. I. & Sunyaev, R. A. Black holes in binary systems. Observational appearance. *Astron. Astrophys.* **24,** 337–355 (1973).

48. Belczynski, K. *et al.* On the maximum mass of stellar black holes. *Astrophys. J.* **714,** 1217–1226 (2010).
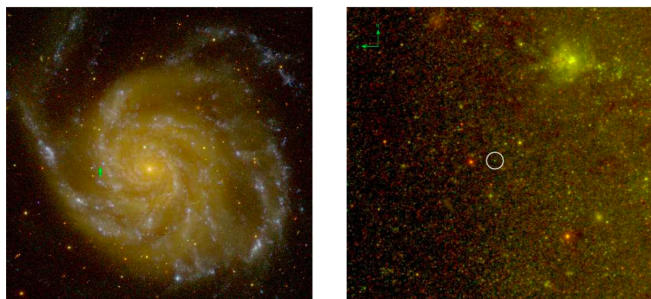
Extended Data Figure 1 | M 101 ULX-1 as observed in the optical region.
Left, M 101 ULX-1 is located on a spiral arm of the face-on grand-design spiral galaxy M 101, as indicated by the arrow. The colour image of M 101 is composed of GALEX NUV, SDSS g, and 2MASS J images. Right, ULX-1 is identified as a blue object with $V = 23.5$ mag at the centre of the 1″ circle on the HST image. The colour image is composed of ACS/WFC F435W, F555W and F814W images.
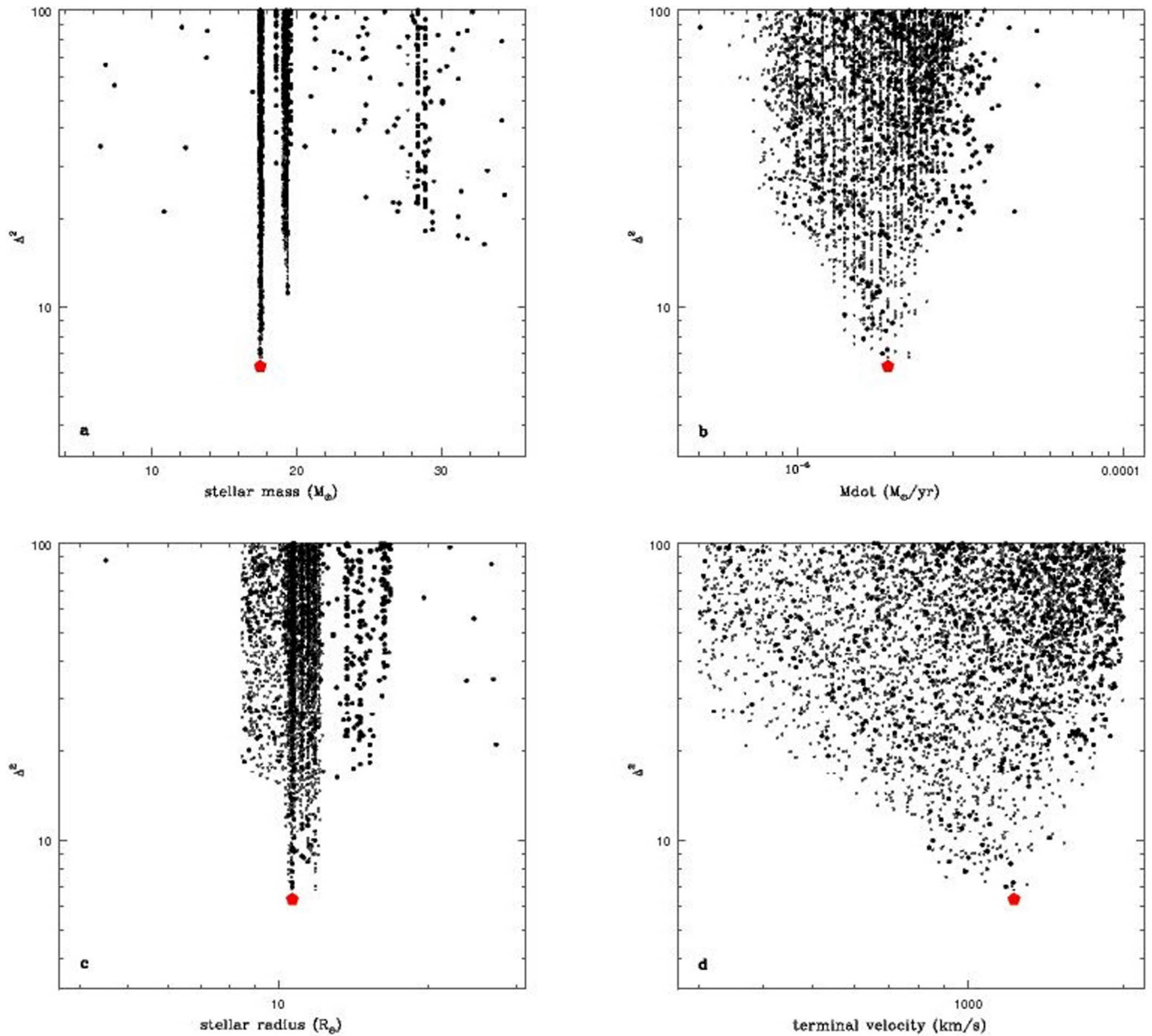
Extended Data Figure 2 | Physical properties of the Wolf-Rayet secondary from spectral line modelling. Distributions of computed Δ^2 as a function of stellar masses (**a**), stellar mass loss rate (**b**), stellar radii (**c**) and terminal velocity (**d**). Here $\Delta^2 = \Sigma_i (EW - EW_i)^2$ computes the difference between observed and synthetic equivalent widths EW for six broad helium lines present in the Gemini/GMOS spectrum. We have computed synthetic spectra for a group of 5,000 real stars from the evolution tracks (as shown by the thick stripes in the mass plot and the radius plot) and for another group of 'fake' stars with continuous distributions in mass, radius and luminosity. The best model is labelled by a filled pentagon in all panels.
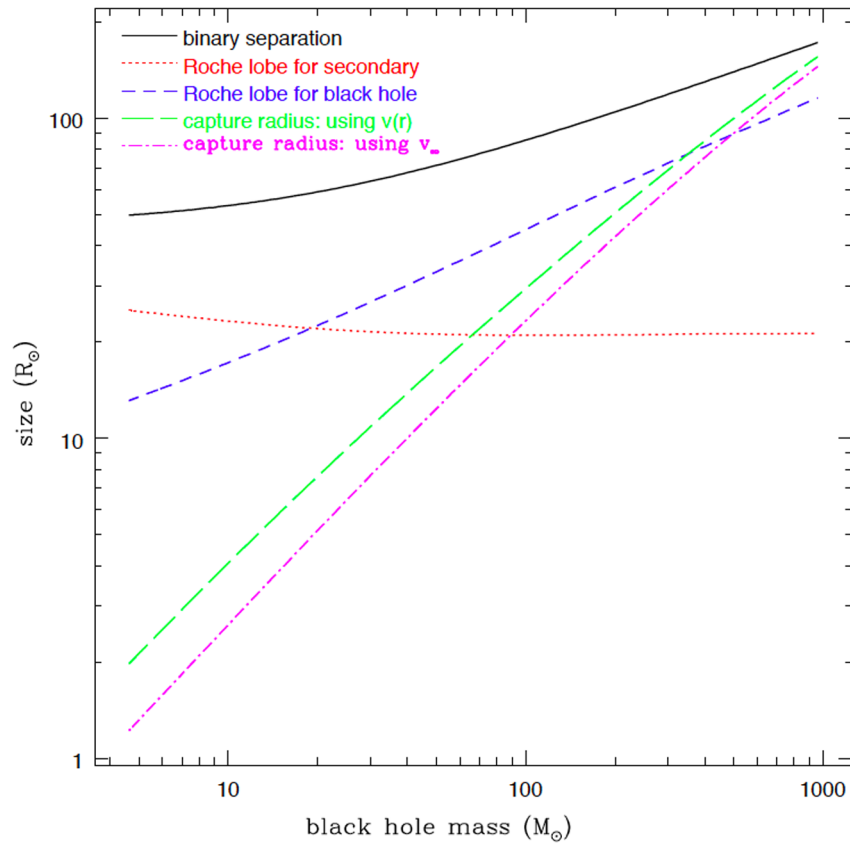
Extended Data Figure 3 | Properties of the Wolf-Rayet/black-hole binary for different black-hole masses. Shown are the binary separation (solid line), the Roche lobe sizes for the Wolf-Rayet star (dotted) and for the black hole (short dashed), the capture radius for the black hole when using the terminal velocity (dash–dotted) or when using a simplified velocity law $v(r) = v_\infty (1 - R_*/r)$ (long dashed).
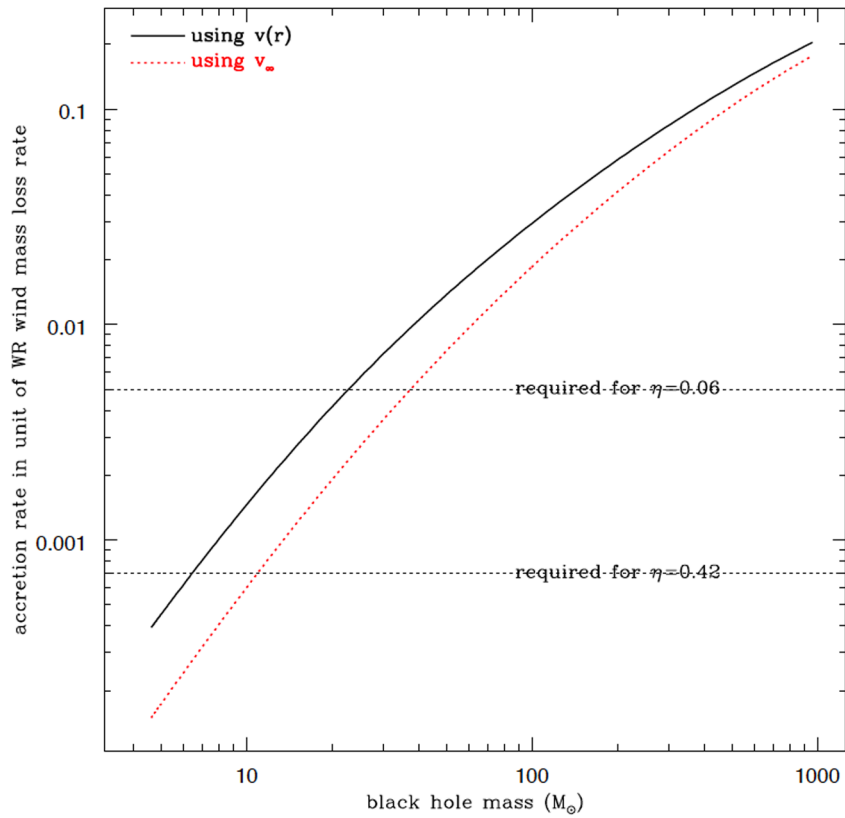
Extended Data Figure 4 | The black-hole accretion rate for different black-hole masses. The accretion rates are computed adopting the terminal velocity (dotted) and a simplified velocity law $v(r) = v_\infty(1 - R_*/r)$ (solid). To power the observed average luminosity of 3×10^{38} erg s^{-1}, the black-hole mass must exceed $13M_\odot$ ($8M_\odot$) using the terminal velocity (the velocity law) for a Kerr black hole ($\eta = 0.42$), and exceed $46M_\odot$ ($28M_\odot$) for a Schwarzschild black hole ($\eta = 0.06$). The two horizontal dotted lines indicate the accretion rates required for $\eta = 0.06$ and $\eta = 0.42$, respectively.
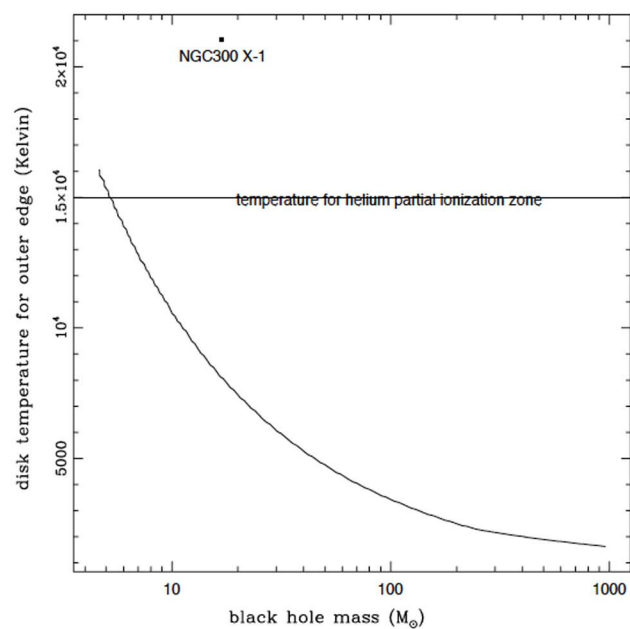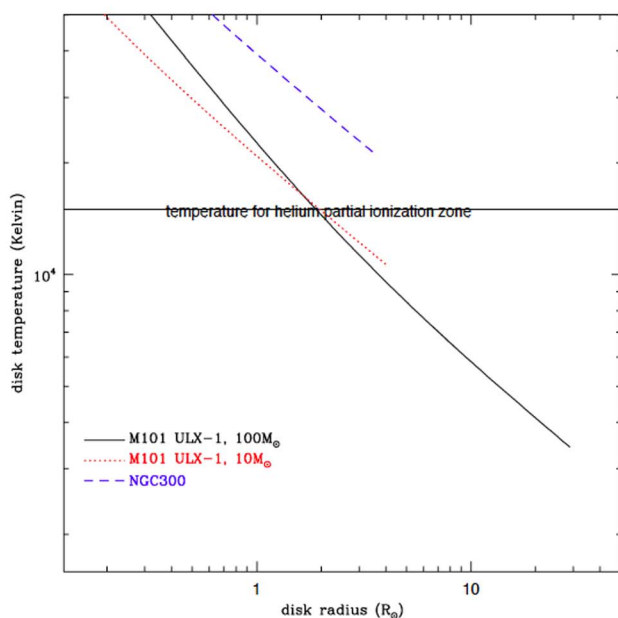
Extended Data Figure 5 | Disk temperature structures for M 101 ULX-1.
a, The disk temperature profiles for M 101 ULX-1 (for $P = 8.24$ days, $M_* = 19M_\odot$, $R_* = 10.7R_\odot$, $M_\bullet = 10M_\odot$ or $100M_\odot$) and NGC300 X-1 (for $P = 32.4$ h $M_* = 26M_\odot$, $R_* = 7.2R_\odot$, $M_\bullet = 16.9M_\odot$; ref 22). **b**, The disk temperature at the outer edge for different black-hole mass in M 101 ULX-1. The horizontal line indicates the temperature required for the helium partial ionization zone.

Extended Data Table 1 | Gemini/GMOS spectroscopic observations of M 101 ULX-1

OBSDATE	MJD	exposure (second)	bary. (km/s)	velocity (km)
2010-02-15	55242.58343	3200	7.4	212
2010-02-16	55243.50615	3200	7.3	236
2010-03-16	55271.54390	3200	0.1	301
2010-03-17	55272.54564	3200	-0.2	—
2010-04-17	55303.47547	4800	-7.7	326
2010-05-13	55329.33126	6400	-12.2	302
2010-05-14	55330.39682	6400	-12.4	256
2010-05-15	55331.37803	6400	-12.5	227
2010-05-18	55334.41410	9600	-13.0	244
2010-05-19	55335.42391	9600	-13.1	305

The columns are: (1) observation date, (2) modified Julian date, (3) exposure time in seconds, (4) barycentric correction computed with rvsao, and (5) the corrected radial velocity as measured with He II 4,686 Å, with an error of 15 km s^{-1} as mainly from the uncertainties in the wavelength calibration.

Extended Data Table 2 | Properties of emission lines

Line ID	FWHM (Å)	E.W. (Å)	Lum. 10^{34}erg/s	model (Å)
HeII 4686	19.3	21.83 ± 0.20	43	21.75
HeI 5876	19.0	34.78 ± 0.29	49	34.21
HeI 6679	18.8	25.74 ± 0.37	24	26.56
HeII 5411	20.5	5.46 ± 0.13	8.3	6.10
HeI 4922	13.4	5.80 ± 0.64	8.4	3.91
HeI 4471	12.1	3.86 ± 0.65	7.0	5.18
H_γ	3.6	1.35 ± 0.22	2.7	
H_β	4.5	7.51 ± 0.06	12	
H_α	4.7	26.54 ± 0.46	34	
$[OIII]$ 4960	4.4	23.70 ± 0.49	40	
$[NII]$ 6548	3.8	3.85 ± 0.39	4.7	
$[NII]$ 6583	4.7	16.66 ± 0.08	18	
$[SII]$ 6716	4.0	4.58 ± 0.07	4.0	
$[SII]$ 6731	4.6	3.81 ± 0.06	3.1	

The columns are: (1) emission line ID, (2) FWHM as obtained from Gaussian fit, which equals 2.35σ, (3) equivalent width, (4) line luminosity in units of 10^{34} erg s^{-1}, and (5) equivalent width from the best Wolf-Rayet synthetic model.

LETTER

Potential for spin–based information processing in a thin–film molecular semiconductor

Marc Warner[1]†, Salahud Din[2], Igor S. Tupitsyn[3], Gavin W. Morley[1]†, A. Marshall Stoneham[1]‡, Jules A. Gardener[1]†, Zhenlin Wu[2], Andrew J. Fisher[1], Sandrine Heutz[2], Christopher W. M. Kay[4] & Gabriel Aeppli[1]

Organic semiconductors are studied intensively for applications in electronics and optics[1], and even spin-based information technology, or spintronics[2]. Fundamental quantities in spintronics are the population relaxation time (T_1) and the phase memory time (T_2): T_1 measures the lifetime of a classical bit, in this case embodied by a spin oriented either parallel or antiparallel to an external magnetic field, and T_2 measures the corresponding lifetime of a quantum bit, encoded in the phase of the quantum state. Here we establish that these times are surprisingly long for a common, low-cost and chemically modifiable organic semiconductor, the blue pigment copper phthalocyanine[3], in easily processed thin-film form of the type used for device fabrication. At 5 K, a temperature reachable using inexpensive closed-cycle refrigerators, T_1 and T_2 are respectively 59 ms and 2.6 µs, and at 80 K, which is just above the boiling point of liquid nitrogen, they are respectively 10 µs and 1 µs, demonstrating that the performance of thin-film copper phthalocyanine is superior to that of single-molecule magnets over the same temperature range[4]. T_2 is more than two orders of magnitude greater than the duration of the spin manipulation pulses, which suggests that copper phthalocyanine holds promise for quantum information processing, and the long T_1 indicates possibilities for medium-term storage of classical bits in all-organic devices on plastic substrates.

The drive to develop spintronics, through precise control and read-out of electron spins, has provided impetus for both fundamental discoveries and practical devices. Whereas initial studies mainly considered solid-state inorganic materials, recent work has focused on more exotic species, in particular single-molecule magnets[5–8]. These tend to be large, complex molecules possessing many electron spins and magnetic nuclei that induce decoherence. Thus, the longest decoherence times are measured at ultralow temperatures with the single-molecule magnets isolated from each other by dilution into either diamagnetic isomorphous host crystals or in frozen solution[4,9]. The latter approach has also been used for other molecular materials, such as N@C$_{60}$ (atomic nitrogen inside a 60-atom carbon cage; ref. 10), that have intrinsically long decoherence times at ambient temperatures.

However, when considering compatibility with current thin-film-based plastic electronic and optoelectronic technologies, and reliability of manufacturing and usage, the potential of simpler molecules, such as copper phthalocyanine (CuPc; Fig. 1a), that can be produced on an industrial scale and readily processed in thin films both for solar energy[3] and molecular electronics[11] should be explored. Moreover, spin-bearing organic molecules often have low spin–orbit coupling, possess a large Hilbert space with many non-degenerate transitions, and can be customized by chemical modification. These positive attributes have led to newer research assessing the potential of macrocycle materials for both spintronics and quantum information processing[5,6,8]. Here we demonstrate that the decoherence times of CuPc are comparable or superior to those of the best molecular systems and can be maintained even in a device-like film configuration on a readily available plastic substrate, Kapton. We achieve this through organic-molecular-beam deposition, co-depositing CuPc with the structurally isomorphous but diamagnetic free base phthalocyanine (H$_2$Pc), allowing the spin-carrying CuPc molecules to be spatially separated while still adopting a well-defined crystal α-phase[12]. Co-deposition reduces spin–spin interactions and therefore decreases the decoherence rates in the ensemble, the measurement of which is always performed as the first test of utility for quantum information processing[4,5,13–17]. To constrain the orientation of the CuPc molecules, and thereby reduce the spectral variation due to the powder averaging of the anisotropic CuPc g factor (g), we deposited the 400-nm-thick CuPc:H$_2$Pc films onto a layer of perylene-3,4,9,20-tetracarboxylic dianhydride. This forces the CuPc and H$_2$Pc molecules to lie nearly flat on the Kapton[18], with the normal to the molecular plane almost perpendicular to the surface (Fig. 1b).

Figure 1c shows the echo-detected field sweeps (EDFSs) of CuPc thin films for different copper spin concentrations. The EDFS is a measurement of the Hahn echo as a function of applied magnetic field[19]. The broadening of spectral features at higher CuPc concentrations results from the increased electronic dipolar interaction. The peak at approximately 325 mT is due to radicals in the Kapton film and oxygen-centred radicals in H$_2$Pc (ref. 20).

Figure 1d is a schematic of the energy levels that give rise to the EDFS spectra for a single molecular orientation, with the normal to the molecular plane parallel to the applied field, as in our measurements. These are simulated in EASYSPIN[21] using the Hamiltonian $H = g\mu_B\boldsymbol{BS} + \sum \boldsymbol{IAS}$ (see Methods for details), the two terms of which respectively represent the Zeeman energy for the electrons within the external field $\boldsymbol{B}$ (μ_B, Bohr magneton) and the sum of the various hyperfine interactions[19]. Copper(II) complexes have been studied extensively[22]: for CuPc the electronic spin is $S = 1/2$ and for both naturally occurring copper isotopes (^{63}Cu and ^{65}Cu) the nuclear spin is $I = 3/2$. The hyperfine coupling of ^{63}Cu is defined by the diagonal matrix $\boldsymbol{A}$ with $A_{xx} = A_{yy} = -83$ MHz and $A_{zz} = -648$ MHz in the molecular frame[19] (these values scale for ^{65}Cu according to the ratio between gyromagnetic ratios). The predominant (>99%) naturally occurring nitrogen isotope (^{14}N) has $I = 1$ and the four nearest-neighbour nitrogens have a hyperfine coupling to the d^9 Cu^{2+} of $A_{xx} = 57$ MHz and $A_{yy} = A_{zz} = 45$ MHz (ref. 23). The red arrows in Fig. 1d indicate the allowed transitions, which, as indicated in the first magnified view, cluster into four groups (owing to the interaction with the spin-3/2 copper nuclei) of nine transitions (owing to the four identical spin-1 nitrogen nuclei). The second magnified view shows the expected intensity variation of the transitions (1:4:10:16:19:16:10:4:1).

The decoherence time, T_2, is the time over which a quantum bit, or qubit, can reliably store quantum information[19], and is obtained by incrementally increasing τ, the time difference between π/2- and π-pulses, in the Hahn echo sequence (Fig. 2b, inset) and measuring the decay of the

[1]London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, London WC1H 0AH, UK. [2]London Centre for Nanotechnology and Department of Materials, Imperial College London, London SW7 2AZ, UK. [3]Pacific Institute of Theoretical Physics, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada. [4]Institute of Structural & Molecular Biology and London Centre for Nanotechnology, University College London, London WC1E 6BT, UK. †Present addresses: Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA (M.W.); Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK (G.W.M.); RMD Inc., 44 Hunt Street, Watertown, Massachusetts 02472, USA (J.A.G.).
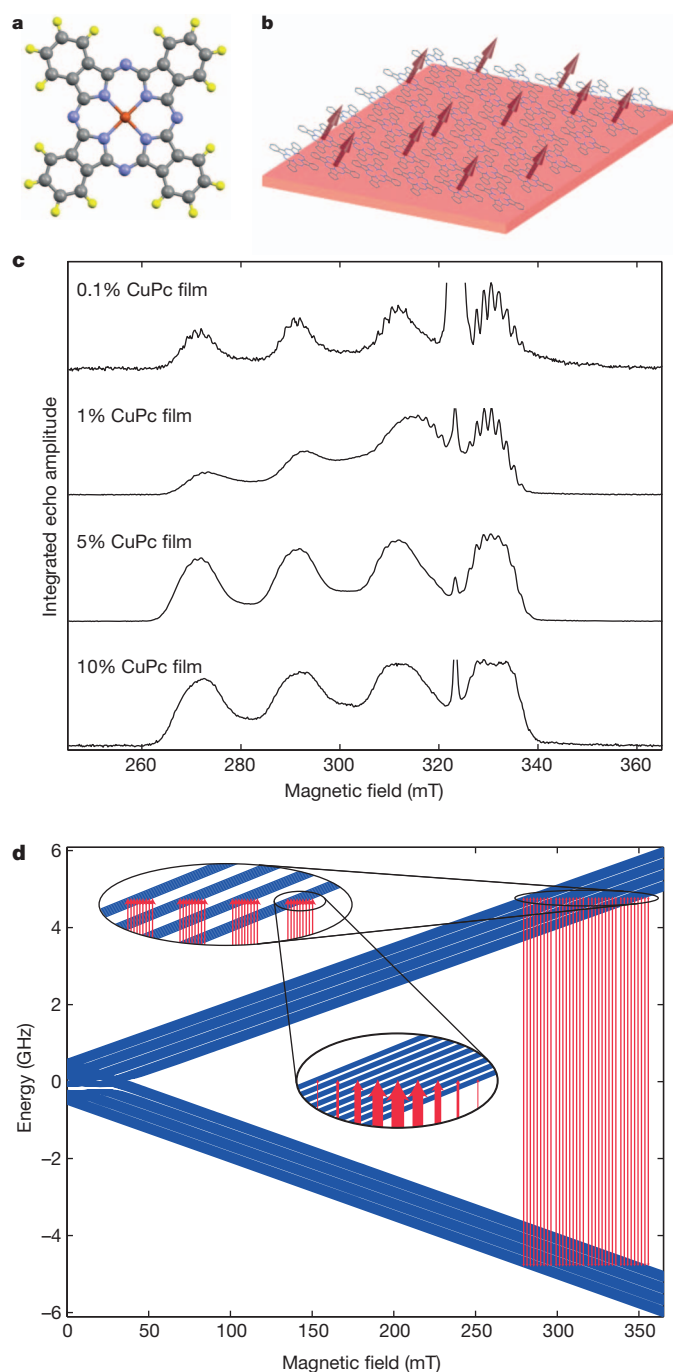‡Deceased.

Figure 1 | Copper phthalocyanine films. a, Copper phthalocyanine molecule, containing copper (red), nitrogen (blue), carbon (grey) and hydrogen (yellow) atoms. **b**, Schematic representation of the dilute film, with dispersed electron spins depicted as arrows. For clarity, only a single molecular layer is shown; films used in experiments are 400 nm, or approximately 1,200 layers, thick. **c**, EDFSs collected at 5 K for different copper concentrations. **d**, Energy level structure of the spin Hamiltonian for a single CuPc molecule oriented perpendicular to the magnetic field, with two magnified views showing the copper hyperfine coupling and the nitrogen hyperfine coupling, respectively.

integrated echo[19]. Figure 2a shows the echo decay for varying concentrations of CuPc. The oscillations at early times in the 0.1% CuPc signal are electron spin-echo envelope modulation oscillations caused by coupling to nitrogen nuclear spins[19]. The inset in Fig. 2a shows a comparison of two EDFSs performed at $\tau = 0.4$ and 2.3 µs, respectively, confirming that the echo at long times is from CuPc and not from another spin-carrying defect.

Because the samples are randomly spin-diluted, the local environments of each spin will differ, leading to a distribution of decoherence times, with no single characteristic T_1 or T_2 for a particular concentration. To extract meaningful times we fit the echo decay with $(\alpha + \beta\sin(\omega t + d))\exp(-t/T_2)$ and plot the decay constant as a function of concentration (Fig. 2b). The fit is chosen to allow the extraction of the underlying T_2, the constant that characterizes the exponentially decaying envelope, in the presence of the sinusoidal electron spin-echo envelope modulation oscillations. The dilute limit provides a measure of the intrinsic T_2 of isolated Cu spins for the α-phase samples grown according to this method; different crystal structures can yield different decoherence times because the distances from the copper atoms to electron and nuclear spins that cause relaxation are modified.

Decoherence in electron spin-echo experiments manifests itself through dephasing. For convenience, we introduce a dimensionless decoherence rate, $\gamma = 2\hbar/T_2\Delta_Q$, that is inversely proportional to the number of easily visible coherent oscillations[5]; $\Delta_Q/\hbar = \omega_{EPR} = 9.71$ GHz is the energy gap between the Cu^{2+} $S = 1/2$ qubit states. At very low concentrations, c, of CuPc in H_2Pc, decoherence arises mainly from the local nuclear spins whose motion dephases the dynamics of the electron spins via the hyperfine interaction. At large concentrations, dephasing is dominated by the pairwise dipolar flip-flop processes between the electronic Cu^{2+} spins. For more details, see Methods.

From the fine structure of echo lines in Fig. 1c in the low-c limit, we see that the peaks, corresponding to the four nearest-neighbour nitrogens, strongly overlap. This means that the nuclear spin polarization diffusion is not confined to an individual nitrogen peak and can 'scan' through the entire multiplet of the nitrogen nuclear spin states coupled to the electronic spin of copper (that is, through any one of the four multiplets in the upper left inset in Fig. 1d). For the nuclear spin bath[5], $\gamma_n = 2(E_n/\Delta_Q)^2$, where E_n is the half-width of the above-mentioned multiplet (the limit $E_n \ll \Delta_Q$ is assumed). Knowing the hyperfine couplings and positions of all the nuclear spins in CuPc (ref. 12), we can compute E_n and, correspondingly, the nuclear-induced contribution, $2\hbar/\gamma_n\Delta_Q$, to the total T_2 of the electronic spin. This contribution is represented in Fig. 2b by the green dashed line. The calculated value of E_n ($\sim$1.25 mK $\ll \Delta_Q$) and the value extracted from Fig. 1c at $c = 0.1\%$ agree with each other, and the dominant contribution comes from the four nearest-neighbour nitrogens. Our calculations yield a nuclear-induced dephasing time of 2.2 µs, which is close to the value, $T_2 = 2.6$ µs, that is observed at $c \lesssim 0.1\%$. At larger concentrations, the dipolar inter-Cu^{2+} processes start to contribute. For $c \lesssim 10\%$, the electronic spin system is in the dilute limit and the electronic dipolar contribution, E_d, to the echo line half-width is also very small compared with the gap Δ_Q. This allows us to write the electronic decoherence rate, γ_d, in the same form as the nuclear rate, but with E_n replaced by E_d. To calculate E_d, we place CuPc molecules on random sites of a 50-nm-diameter crystalline granule to achieve the desired concentration, and fill the remaining sites with H_2Pc molecules (Methods). Then, after configurational averaging over the CuPc positions, we obtain the electronic pairwise-induced contribution to the total T_2 (Fig. 2b, blue dashed line). The solid red curve shows the total theoretical value, $T_2 = 2\hbar/\Delta_Q(\gamma_n + \gamma_d)$, which we compare with the experimental measurements (blue circles).

To establish experimentally the magnetic field dependence of T_2, we used only diluted CuPc powders, because the smaller resonant cavity of a Q-band spectrometer is not amenable to stacking films. The inset in Fig. 2b demonstrates that in the more dilute α-phase CuPc powders, T_2 is weakly dependent on field, at least up to 1 T.

A lower bound on the population relaxation time, T_1, was measured with an inversion recovery sequence (Fig. 2d, inset) in which t was varied and the integrated echo intensity was monitored[23]. Although this measurement does not strictly include only T_1 effects—there are additional contributions to the echo decay from spectral diffusion—it does reflect the real lifetime of the classical state of the qubit.

Once again, the distribution of environments leads to a signal consisting of a superposition of many exponential decays, with no single
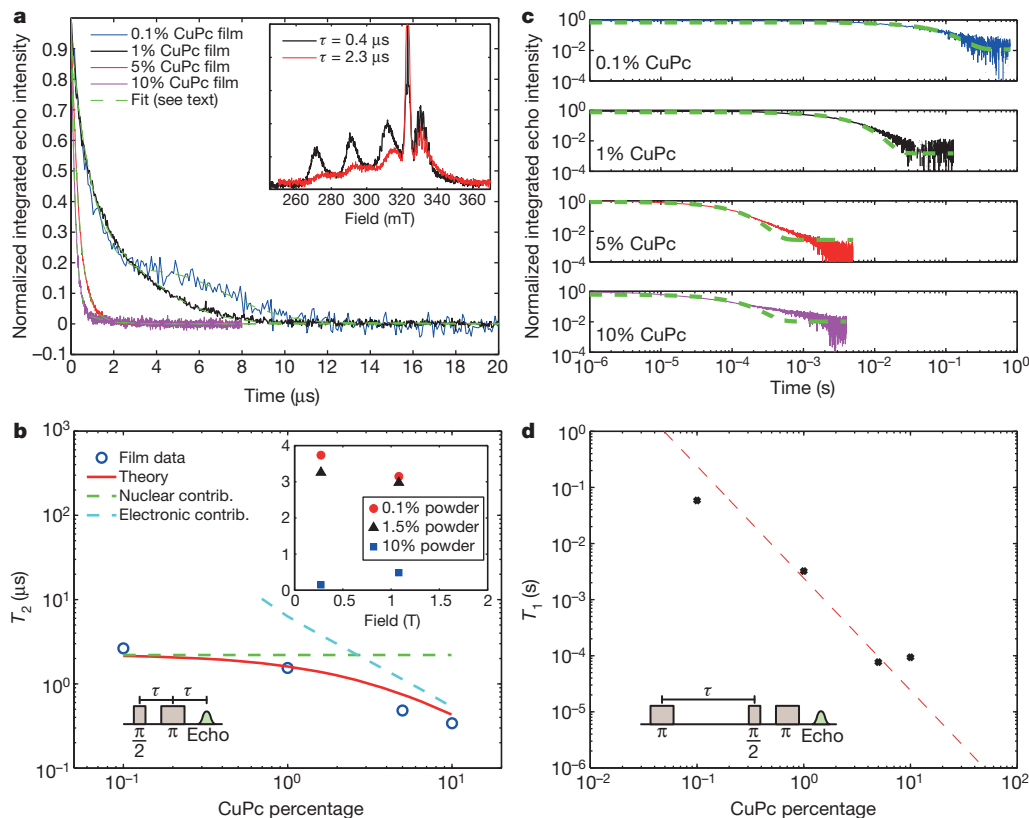
Figure 2 | Concentration dependence of decoherence and relaxation times.
a, Echo decay with time difference between $\pi/2$- and π-pulses for different CuPc concentrations, with fits (see text for details), recorded at 5 K with a 311.5-mT magnetic field. Inset, EDFSs at short and long times showing that the signal is from the CuPc molecules, normalized to the Kapton peak. The differences in shape of the two EDFS spectra indicate a slight field dependence of the relaxation rate. **b**, Experimental T_2 times, extracted from fits, plotted together with theoretical ones as a function of concentration. Lower inset, Hahn echo

pulse sequence. Upper inset, respective T_2 values of $(CuPc)_p(H_2Pc)_{1-p}$ powders at 0.273 and 1.08 T, demonstrating the minor field dependence between these values; p is a nominal composition derived from the ratio of starting materials. **c**, Fitted decays (see text for details) of the Hahn echo as part of an inversion recovery sequence recorded at 5 K for different CuPc concentrations. **d**, T_1 values extracted from **c** (symbols), with fit (line; see text for details). Inset, inversion recovery pulse sequence.
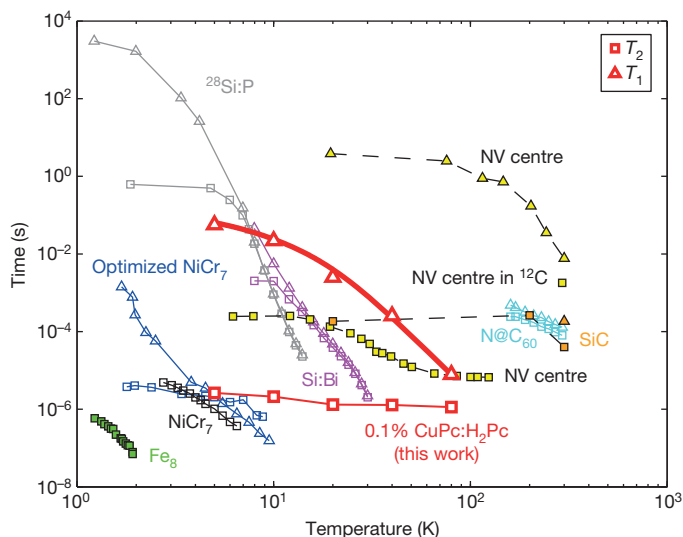
Figure 3 | Temperature dependence of decoherence and relaxation times.
T_1 and T_2 versus temperature for 0.1% CuPc:H$_2$Pc film measured in a 311.5-mT magnetic field, and other common spin qubit systems. It should be noted that the data in this comparison were not all collected at the same magnetic field, but represent the best available comparison from the literature. Data are from ref. 4 (NiCr$_7$), ref. 5 (Fe$_8$), ref. 9 (optimized NiCr$_7$), ref. 10 (N@C$_{60}$ in CS$_2$), ref. 13 (^{28}Si:P), ref. 14 (Si:Bi), ref. 15 (nitrogen–vacancy (NV) centre) and ref. 16 (nitrogen–vacancy centre in ^{12}C) and ref. 17 (defect in SiC). Lines between points are plotted as guides to the eye, except for that for the T_1 of CuPc:H$_2$Pc, which was fitted as described in the text.

characteristic decay constant (Fig. 2c). However, by fitting to a single exponential, we can extract a limiting case (for stretched-exponential fits see Extended Data Figs 1, 2 and 3). The fits are imperfect at short times and high concentrations, but improve with dilution, because the probability of the molecules experiencing a similar, isolated, environment increases. The characteristic times derived from the fits at $T = 5$ K are plotted in Fig. 2d, and are approximately proportional to $1/c^2$. Inspection of the decay curves and the corresponding fits shows that these are the shortest times in a distribution of relaxation times; there are also longer times, which become more obvious for higher concentrations. This is expected for the depolarization associated with the non-secular (and, hence, non-magnetization-conserving) part of the dipole–dipole interaction[24], for which the characteristic matrix elements scale as r^{-3} and, hence, as c; the relaxation rate in second-order perturbation theory is therefore proportional to c^2.

To establish whether our films might be useful at or above the boiling point of liquid nitrogen, we measured the temperature dependences of T_1 and T_2. Figure 3 shows these for a 0.1% CuPc film. Experiments, such as ours, that look at the longest decay times naturally select subpopulations of isolated spins. As the temperature is raised, the echo decay becomes closer in character to a mono-exponential decay, indicating—as expected—that differences between spin environments are being averaged away.

In Fig. 3, we also include T_2 (and, where available, T_1) results for a selection of comparable spin qubit candidates in the sense that they use the same control mechanism. Over the easily accessed 5–80 K temperature range, the CuPc:H$_2$Pc films are superior to all other molecular options, with the exception of N@C$_{60}$ solvated in CS$_2$ (which makes it

a

b

c
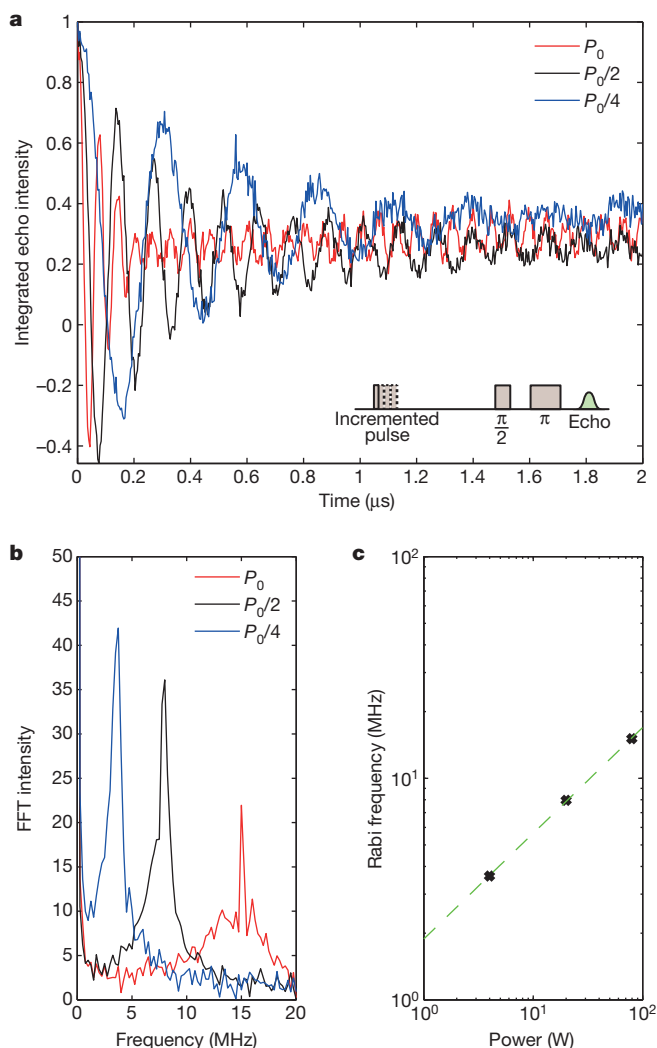
Figure 4 | **Controlling a single qubit. a**, Rabi oscillations of a 0.1% $CuPc:H_2Pc$ film recorded at 5 K and 330.5 mT for different microwave powers. **b**, Fast Fourier transform (FFT) of the Rabi oscillations. **c**, Rabi frequency (symbols) is proportional (line; see text for details) to the microwave field strength, which scales with the square root of the power.

unsuitable for incorporation into devices). Indeed, sufficiently isolated CuPc molecules have decoherence times of approximately 1 μs, even at 80 K, which makes the molecular thin film superior to silicon doped even with the favourable heavy group V element bismuth at temperatures above 30 K (ref. 14). The nitrogen–vacancy centres in diamond have longer relaxation times up to room temperature, but are much more challenging to introduce into the appropriate host in a controlled way. Furthermore, the difference between the T_2 and the long T_1 of the CuPc films demonstrates the potential for optimization, because T_1 provides an upper bound on T_2.

To quantify the T_1 relaxation mechanisms, we fitted the temperature dependence to the form $1/T_1 = a + b/(\exp(-\Delta/T) - 1)$, implying two processes: a temperature-independent spin–spin interaction, where $a = 17.8\ s^{-1}$, and an Orbach process[25]. The energy scale of the Orbach process, Δ, is specified to be 69 K in accordance with literature values for optical phonon excitations in the system[26], and b is then found to be $2.6 \times 10^4\ s^{-1}$. The phonon density of states relevant for decoherence in the weakly bonded stacks of rigid molecules is not expected to be Debye-like; instead, it will be dominated by optical modes associated with motion of the copper atoms relative to their aromatic hosting rings. This fit corroborates the concentration dependence in Fig. 2, because both suggest that T_1 is largely determined by spin–spin interactions at

5 K. The fit could be refined with additional terms[25], but the effects of other processes are not sufficiently distinguished by the data to allow conclusive identification.

Rabi oscillations, that is, the coherent driving of electrons between the two Zeeman-split energy levels, exemplify the ability to manipulate spins both for spintronics and quantum information processing. In Fig. 4a, we show how Rabi oscillations depend on microwave power. The capacity to rotate the qubits arbitrarily to any point on the Bloch sphere is one of the two main requirements for creating universal quantum gates, the other being the ability to entangle qubits. Figure 4c shows that the frequencies of the oscillations, found by Fourier transforming the signal (Fig. 4b), scale linearly with microwave amplitude, which in turn scales with the square root of the power[18]. The decay of the Rabi oscillations is determined by T_2^*, which includes both homogeneous and inhomogeneous broadening processes[19], and accordingly includes additional dephasing due to the inhomogeneous magnetic field that each spin encounters. It is therefore shorter than T_2; in this case, $T_2^* = 0.4$ μs at 5 K.

Through further sample optimization, it is likely that the decoherence and relaxation times could be extended. Indeed, our theoretical studies suggest that nitrogen-isotope enrichment is preferable to deuteration here. However, the values of T_1 and T_2 that we measure are already greater than those for single-molecule magnets, which require non-trivial synthesis. For example, T_2 was found to be ~1 μs in optimized Cr_7Ni at 5 K (ref. 9), even with the additional complication of deuteration of both molecule and solvent. In addition, the thin-film samples described here are in a form suitable for device processing[11], can be prepared by a range of deposition techniques[18,27], have the added advantage of being structurally flexible, and are both chemically and thermally robust. The further freedoms to dilute with non-magnetic analogues, to use different chemical motifs to control non-magnetic properties[28], to tune magnetic interactions with minor structural alterations[27,29], to excite at optical frequencies[3] and to inject spins into CuPc with high efficiency[30] make these systems attractive alternatives even to conventional inorganic semiconductors, for which exceptional care must be taken to produce truly random mixtures of magnetic and non-magnetic atoms.

METHODS SUMMARY

To prepare the films, we first grew a 20-nm layer of perylene-3,4,9,20-tetracarboxylic dianhydride on flexible, 25-μm-thick Kapton substrates at $0.2\ Å\ s^{-1}$ by organic-molecular-beam deposition in a Kurt J. Lesker SPECTROS system with a base pressure of around 5×10^{-7} mbar. On this layer, and without breaking vacuum, 400 nm of $CuPc:H_2Pc$ was grown by co-deposition of CuPc and H_2Pc from two individual Knudsen cells. The concentration, c, of the films is expressed as the percentage of CuPc relative to H_2Pc by mass; the ratio, r, of numbers of molecules—which is useful for calculating distances between spins—is then the product of the concentration and the ratio of H_2Pc mass to CuPc mass, or $0.893c$. All chemicals were purchased from Sigma-Aldrich and purified using two cycles of gradient sublimation.

Films of 0.1% CuPc of area ~9 cm^2, 1% CuPc of area ~7 cm^2, 5% CuPc of area ~6 cm^2 and 10% CuPc of area ~5 cm^2 were sliced and packed into Suprasil EPR tubes. Each tube was aligned with the magnetic field by orienting the sample to maximize the EDFS width, and the X-band (9-GHz) microwave response was measured using a Bruker Elexsys E580 pulsed EPR spectrometer with a Bruker ER 4118X-MD5 dielectric ring resonator. Pulsed Q-band (34-GHz) measurements were performed on the same spectrometer using the 580U-FTQ accessory and an EN 5107D2 Q-band resonator (both from Bruker). The temperature was controlled using an Oxford Instruments CF935 cryostat, allowing studies in the range 4–300 K.

1. Sirringhaus, H. *et al.* Two-dimensional charge transport in self-organized, high-mobility conjugated polymers. *Nature* **401,** 685–688 (1999).

2. Dediu, V. A., Hueso, L. E., Bergenti, I. & Taliani, C. Spin routes in organic semiconductors. *Nature Mater.* **8,** 707–716 (2009).
3. Lobbert, G. Phthalocyanines in *Ullmann's Encyclopedia of Industrial Chemistry* (Wiley-VCH, 2000).
4. Ardavan, A. *et al.* Will spin-relaxation times in molecular magnets permit quantum information processing? *Phys. Rev. Lett.* **98,** 057201 (2007).
5. Takahashi, S. *et al.* Decoherence in crystals of quantum molecular magnets. *Nature* **476,** 76–79 (2011).
6. Bertaina, S. *et al.* Quantum oscillations in a molecular magnet. *Nature* **453,** 203–206 (2008).
7. Ishikawa, N. Single molecule magnet with single lanthanide ion. *Polyhedron* **26,** 2147–2153 (2007).
8. Hill, S., Edwards, R. S., Aliaga-Alcalde, N. & Christou, G. Quantum coherence in an exchange-coupled dimer of single-molecule magnets. *Science* **302,** 1015–1018 (2003).
9. Wedge, C. J. *et al.* Chemical engineering of molecular qubits. *Phys. Rev. Lett.* **108,** 107204 (2012).
10. Morton, J. J. L. *et al.* Electron spin relaxation of $N@C_{60}$ in CS. *J. Chem. Phys.* **124,** 014508 (2006).
11. Bao, Z., Lovinger, A. J. & Dodabalapur, A. Organic field-effect transistors with high mobility based on copper phthalocyanine. *Appl. Phys. Lett.* **69,** 3066–3068 (1996).
12. Hoshino, A., Takenaka, Y. & Miyaji, H. Redetermination of the crystal structure of alpha-copper phthalocyanine grown on KCl. *Acta Crystallogr. B* **59,** 393–403 (2003).
13. Tyryshkin, A. M. *et al.* Electron spin coherence exceeding seconds in high-purity silicon. *Nature Mater.* **11,** 143–147 (2012).
14. Morley, G. *et al.* The initialization and manipulation of quantum information stored in silicon by bismuth dopants. *Nature Mater.* **9,** 725–729 (2010).
15. Takahashi, S., Hanson, R., van Tol, J., Sherwin, M. S. & Awschalom, D. D. Quenching spin decoherence in diamond through spin bath polarization. *Phys. Rev. Lett.* **101,** 047601 (2008).
16. Balasubramanian, G. *et al.* Ultralong spin coherence time in isotopically engineered diamond. *Nature Mater.* **8,** 383–387 (2009).
17. Koehl, W. F., Buckley, B. B., Heremans, F. J., Calusine, G. & Awschalom, D. D. Room temperature coherent control of defect spin qubits in silicon carbide. *Nature* **479,** 84–87 (2011).
18. Heutz, S., Cloots, R. & Jones, T. Structural templating effects in molecular heterostructures grown by organic molecular-beam deposition. *Appl. Phys. Lett.* **77,** 3938–3940 (2000).
19. Schweiger, A. & Jeschke, G. *Principles of Pulse Electron Paramagnetic Resonance* (Oxford Univ. Press, 2001).
20. Assour, J. & Harrison, S. On origin of unpaired electrons in metal-free phthalocyanine. *J. Phys. Chem.* **68,** 872–876 (1964).
21. Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **178,** 42–55 (2006).
22. Stoneham, A. M. The theory of spin relaxation of copper in a Tutton salt crystal. *Proc. Phys. Soc.* **85,** 107 (1965).
23. Finazzo, C., Calle, C., Stoll, S., Van Doorslaer, S. & Schweiger, A. Matrix effects on copper(ii)phthalocyanine complexes: a combined continuous wave and pulse EPR and DFT study. *Phys. Chem. Chem. Phys.* **8,** 1942–1953 (2006).
24. Van Vleck, J. The dipolar broadening of magnetic resonance lines in crystals. *Phys. Rev.* **74,** 1168–1183 (1948).
25. Abragam, A. & Bleaney, B. *Electronic Paramagnetic Resonance of Transition Metal Ions* (Oxford Univ. Press, 1970).
26. Szybowicz, M. *et al.* High temperature study of FT-IR and Raman scattering spectra of vacuum deposited CuPc thin films. *J. Mol. Struct.* **704,** 107–113 (2004).
27. Wang, H. *et al.* Ultralong copper phthalocyanine nanowires with new crystal structure and broad optical absorption. *ACS Nano* **4,** 3921–3926 (2010).
28. Opitz, A., Wagner, J., Brutting, W., Hinderhofer, A. & Schreiber, F. Molecular semiconductor blends: microstructure, charge carrier transport and application in photovoltaic cells. *Phys. Status Solidi A* **206,** 2683–2694 (2009).
29. Heutz, S. *et al.* Molecular thin films: a new type of magnetic switch. *Adv. Mater.* **19,** 3618–3622 (2007).
30. Cinchetti, M. *et al.* Determination of spin injection and transport in a ferromagnet/organic semiconductor heterojunction by two-photon photoemission. *Nature Mater.* **8,** 115–119 (2009).

Author Contributions M.W. conducted the electron spin resonance measurements with input and supervision from G.A. and C.W.M.K. S.D., J.A.G. and Z.W. made and characterized the samples with input and supervision from S.H.. M.W., G.W.M., A.M.S., A.J.F., C.W.M.K. and G.A. analysed data, I.S.T. performed theoretical work, and M.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W. (marc.warner@ucl.ac.uk) or G.A. (gabriel.aeppli@ucl.ac.uk).

METHODS

Experimental methods. To prepare the films, a 20-nm layer of perylene-3,4,9,20-tetracarboxylic dianhydride was grown on flexible, 25-μm-thick Kapton substrates at $0.2\,\text{Å s}^{-1}$ by organic-molecular-beam deposition in a Kurt J. Lesker SPECTROS system with a base pressure of around 5×10^{-7} mbar. On this, and without breaking vacuum, 400 nm of CuPc:H_2Pc was grown by co-deposition from two individual Knudsen cells. The total rate of deposition onto the substrate was maintained at $1\,\text{Å s}^{-1}$, and the individual rates of CuPc and H_2Pc deposition adjusted depending on the desired stoichiometry of the films. The rates were controlled with quartz crystal monitors placed at the sources and at the substrate. The concentration of the films is expressed as the percentage of CuPc relative to H_2Pc by mass. All chemicals were purchased from Sigma-Aldrich and purified using two cycles of gradient sublimation.

Films of 0.1% CuPc of area $\sim 9\,\text{cm}^2$, 1% CuPc of area $\sim 7\,\text{cm}^2$, 5% CuPc of area $\sim 6\,\text{cm}^2$ and 10% CuPc of area $\sim 5\,\text{cm}^2$ were sliced and packed into Suprasil EPR tubes. Each tube was aligned to the magnetic field by orienting the sample to maximize the EDFS width. Powders were prepared as follows. First, 3 ml of concentrated H_2SO$_4$ (98%) and 15 ml of IPA were cooled for 10 min in ice water. Then 0.3 g of CuPc or H_2Pc (Sigma-Aldrich; purity, 97%) was dissolved in the acid at a concentration of $18 \times 10^{-6}\,\text{mol dm}^{-3}$ while stirring continuously. The acid paste was then put drop by drop into the IPA with continuous agitation. The mixture was filtered and washed in distilled water and acetone, and then the precipitate was left to dry on the filter paper for 30 min. The precipitate was finally allowed to dry completely in a desiccator overnight. The yield was estimated to be 50%.

Pulsed X-band (9-GHz) measurements were made with a Bruker Elexsys E580 pulsed EPR spectrometer with a Bruker dielectric ring resonator ER 4118X-MD5. A pulse of 40 ns and a $\pi/2$-pulse of 20 ns were used. Pulsed Q-band (34-GHz) measurements were performed on the same spectrometer using the 580U-FTQ accessory and a EN 5107D2 Q-Band resonator (both from Bruker).

In the Rabi pulse sequence, when the attenuation of the microwaves was changed, the detection sequence power was corrected by adjusting the pulse lengths appropriately. The microwave frequency was approximately 9.71 GHz, and the pulse power is specified by the manufacturer to be 0.3 kW at the attenuation used. The temperature was controlled using a CF935 Cryostat, allowing studies in the range 4–300 K.

Theoretical methods. The Cu^{2+} ion in the $C_{32}H_{16}N_8Cu$ molecule has electronic spin $S = 1/2$ and nuclear spin $I^{Cu} = 3/2$. The nitrogens (^{14}N), hydrogens (^{1}H) and carbons (^{12}C) (the naturally occurring isotopes) are characterized by the nuclear spins $I^N = 1$, $I^H = 1/2$ and $I^C = 0$, respectively. The samples used in the experiments are 400-nm thin films of $\sim$50-nm, nearly spherical CuPc:H_2Pc granules. The H_2Pc molecule has the same nuclear spins but lacks the Cu atom at the centre and is thus nonmagnetic. The CuPc:H_2Pc granules have the α-phase brick-stack lattice structure with the lattice constants $a = 12.9\,\text{Å}$, $b = 3.77\,\text{Å}$ and $c = 12.24\,\text{Å}$ and angles $\alpha = 96.22°$, $\beta = 90.62°$ and $\gamma = 90.32°$ (ref. 12).

The Hamiltonian, describing an ensemble of CuPc molecules in external magnetic field **B**, reads as

$$H = -\mu_B \sum_i \sum_{\mu=x,y,z} g_\mu B_\mu^i S_\mu^i + \sum_{n,i} \sum_{\mu,\nu=x,y,z} A_{\mu\nu}^{ni} I_\mu^n S_\nu^i$$
$$+ \sum_{i<j} \sum_{\mu,\nu=x,y,z} V_{\mu\nu}^{ij} S_\mu^i S_\nu^j \tag{1}$$

where Greek letters indicate space indices, n enumerates nuclear species, and i and j indicate electronic spins. The first and third terms describe the anisotropic electronic Zeeman ($\boldsymbol{B^i}$ is the total field acting on the ith electronic spin) and dipole–dipole interactions between the Cu^{2+} electronic spins in the sample. The second term includes (omitting the electronic spin index) the hyperfine interaction between the electronic and nuclear spins of Cu ($A^{Cu}_{\mu\mu}$); the hyperfine interactions between the electronic spin of copper and the nuclear spins of the four nearest-neighbour nitrogens in one molecule ($A^N_{\mu\mu}$); and the dipole–dipole interactions between the copper electronic spin and all the nuclear spins, both in the local molecule and in the rest of sample ($A^n_{\mu\nu}$). There are other terms in the Hamiltonian, describing internuclear interactions, nuclear quadrupolar terms and so on, but their effects are negligibly small (see below). The components of the anisotropic g factor and the strengths of the hyperfine interactions in CuPc are known[23]: $g_z = 2.1577$, $g_x = g_y = 2.039$; $A^{Cu}_{zz} = -648\,\text{MHz}$, $A^{Cu}_{xx} = A^{Cu}_{yy} = -83\,\text{MHz}$; $A^N_{xx} = 57\,\text{MHz}$, $A^N_{yy} = A^N_{zz} = 45\,\text{MHz}$. The positions of all atoms within CuPc are also known[12]. The hybridization between the electronic orbitals of the copper ion and ligands decreases the effective value of the copper's electronic spin in CuPc. This effect is described by the covalency parameter, α^2, and in CuPc it varies between 0.72 and 0.77 (ref. 23). In our calculations, we use $\alpha^2 = 0.74$.

Decoherence. Because Cu^{2+} has $S = 1/2$, the spin–phonon channel of decoherence is absent and two relevant decoherence mechanisms here are related to dephasing of the electronic spin dynamics due to interactions with the nuclear and electronic spin baths (precession of local nuclear spins around the changing directions of the local field, created by the electronic spin, and flip-flop transitions in pairs of the dipole-coupled electronic spins). The essential parameter is a dimensionless decoherence rate, γ, which is inversely proportional to the number of coherent oscillations before decoherence sets in and is related to the decoherence time by $T_2 = 2\hbar/\gamma\varDelta_Q$, where $\varDelta_Q$ is the energy gap between the electronic spin states. Details on the corresponding mechanisms with application to quantum nanomagnets can be found in refs 5, 31, 32.

Nuclear spin bath. The half-width, E_n, of the Gaussian multiplet of nuclear spin states coupled to the electronic spin of Cu^{2+} is given by

$$E_n^2 = \sum_k \frac{(\omega_k^{||})^2 I_k(I_k+1)}{3}$$

where $\omega^{||}_k$ are the differences between the energies of interaction of the electronic spin and the kth nuclear spin when the electronic spin is in the two different (qubit) states[32,33]. They depend on positions of nuclear spins and hyperfine couplings. With this information, E_n can be calculated for any value of the external magnetic field or, equivalently, for any value of the gap $\varDelta_Q = \hbar\omega_{EPR}$. For a CuPc molecule with naturally occurring isotopes at $\varDelta_Q = 9.71$ GHz, we obtain $E_n \approx 1.25$ mK, with the dominant contribution from the four nearest-neighbour nitrogens. The calculated E_n and the experimentally measured half-width of the echo line at $c = 0.1\%$ (extracted from Fig. 1c) agree with each other. This correspondence indicates that the calculation captures the essential physics, and thus justifies neglecting other terms (internuclear and so on) in the Hamiltonian (equation (1)). In the limit $E_n \ll \varDelta_Q$, the nuclear dimensionless decoherence rate, γ_n, can be obtained perturbatively[32,33], yielding $\gamma_n = 2(E_n/\varDelta_Q)^2$. The nuclear-spin-bath-induced contribution to the electronic decoherence time, T_2, at $k_BT \gg E_n$ is temperature independent. It is also independent of the CuPc concentration.

Electronic spin bath. The electronic dipolar contribution, E_d, to the echo line half-width increases with the concentration of CuPc molecules, and to calculate it we use the Van Vleck method[24]. The electronic decoherence times in this work were measured at $5\,\text{K} < T < 80\,\text{K}$, which is very large compared with all the other parameters. Thus, it is safe to consider the limit of infinite temperatures. In this limit, the generalized expression for the Van Vleck second moment with **B** parallel to $\hat{Z}$ and for an anisotropic g factor is

$$E_d^2 = \frac{1}{N_m} \sum_{i<j} \frac{W_{ij}^2 S(S+1)}{3}$$

where

$$W_{ij} = \frac{1}{2}(2g_{||}^2 + g_\perp^2) \frac{\mu_0}{4\pi} \frac{\mu_B^2}{|r_{ij}|^3} \left(1 - \frac{3z_{ij}^2}{|r_{ij}|^2}\right)$$

and N_m is the number of CuPc molecules, $|r_{ij}|$ is the distance between two molecules and the dipolar sum can be calculated numerically for any sample geometry. We limit our consideration to one 50-nm granule, varying the number of CuPc molecules with concentration and randomly placing them on the lattice. The remaining sites of the α-phase lattice are occupied by nonmagnetic H_2Pc molecules. The configurational averaging (that is, averaging over the positions of the CuPc molecules) corresponds to sampling a large ensemble of granules. Owing to the templating effect, molecules in all the granules are nearly parallel to the Kapton film, which makes the orientational averaging unnecessary. In small α-phase CuPc:H_2Pc granules, the second momentum always remains finite, decreasing together with concentration (no long-tail divergences). At concentrations $c \leq 10\%$, the electronic spin system is in the dilute limit, characterized by $E_d \ll \varDelta_Q$. In this limit, we can follow the case of the nuclear-spin-bath approach and obtain the electronic dimensionless decoherence rate perturbatively in the form $\gamma_d = 2(E_d/\varDelta_Q)^2$. At $k_BT \lesssim \varDelta_Q$, the electronic dipolar-flip-flop-induced contribution to the total electronic T_2 is temperature dependent.

31. Morello, A., Stamp, P. C. E. & Tupitsyn, I. S. Pairwise decoherence in coupled spin qubit networks. *Phys. Rev. Lett.* **97,** 207206 (2006).
32. Stamp, P. C. E. & Tupitsyn, I. S. Crossover in spin-boson and central spin models. *Chem. Phys.* **296,** 281–293 (2004).
33. Stamp, P. C. E. & Tupitsyn, I. S. Coherence window in the dynamics of quantum nanomagnets. *Phys. Rev. B* **69,** 014401 (2004).
34. Barbon, A., Brustolon, M., Lisa Maniero, A., Romanelli, M. & Brunel, L.-C. Dynamics and spin relaxation of tempone in a host crystal. An ENDOR, high field EPR and electron spin echo study. *Phys. Chem. Chem. Phys.* **1,** 4015–4023 (1999).
35. Romanelli, M. & Kevan, L. Evaluation and interpretation of electron spin-echo decay part I: rigid samples. *Concepts Magn. Reson.* **9,** 403–430 (1997).
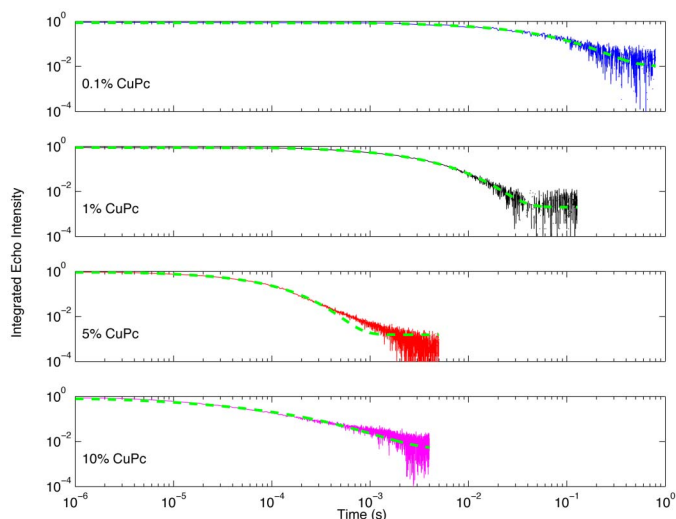
Extended Data Figure 1 | Stretched-exponential T_1 fits. Inversion recovery echoes for varying CuPc concentrations fitted with stretched exponentials. The T_1 decay of each echo magnitude can also be fitted to a stretched exponential, $A\exp(-x/k)^{\beta}$, which is a form characteristic of the random environment that the CuPc molecules experience. In particular, the more isolated molecules will show slower relaxation[34]. However, because the stretched exponential is a phenomenological fit, it must be interpreted with care, particularly in cases where the underlying distribution of relaxation times is highly non-trivial. This is the case in this work, where relaxation times depend strongly on long-range dipolar interactions and, therefore, the finite size of the crystallites[35].
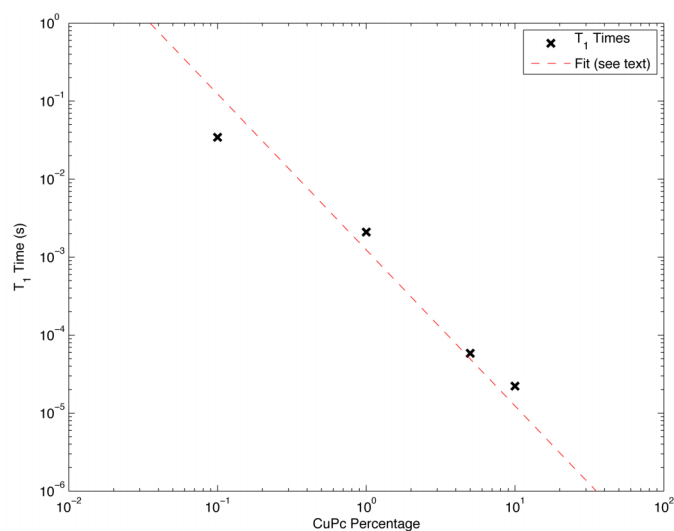
Extended Data Figure 2 │ Decay times of stretched-exponential fits. Decay times extracted from the fits in Extended Data Fig. 1 and plotted against CuPc concentration. The concentration dependence of T_1 is not greatly affected by the change in fit. This allows the interpretation of the data based on the simpler mono-exponential fits (main text).
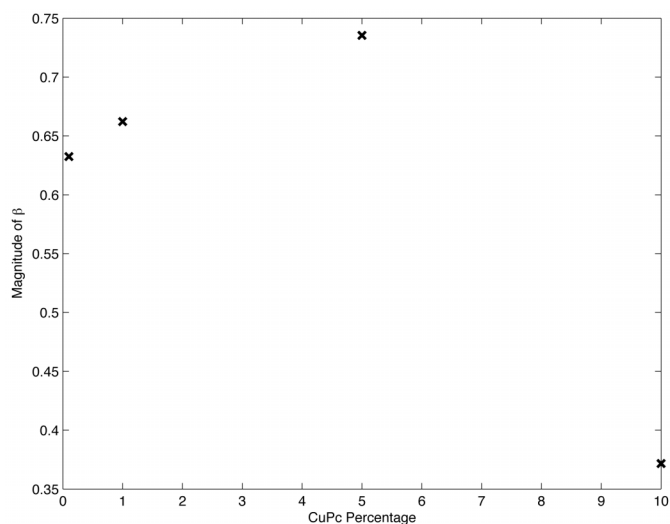
Extended Data Figure 3 | Power-law exponents of stretched-exponential fits. Magnitudes of the power-law exponent, β, in the fits in Extended Data Fig. 2 plotted against CuPc concentration. In a uniform environment, $\beta = 1$ for the population of spins. The greater is the deviation from this value, the larger is the proportion of long-lived isolated spins relative to the average.

LETTER

Perovskite oxides for visible–light–absorbing ferroelectric and photovoltaic materials

Ilya Grinberg[1], D. Vincent West[2], Maria Torres[3], Gaoyang Gou[1], David M. Stein[2], Liyan Wu[2], Guannan Chen[3], Eric M. Gallo[3], Andrew R. Akbashev[3], Peter K. Davies[2], Jonathan E. Spanier[3] & Andrew M. Rappe[1]

Ferroelectrics have recently attracted attention as a candidate class of materials for use in photovoltaic devices, and for the coupling of light absorption with other functional properties[1–7]. In these materials, the strong inversion symmetry breaking that is due to spontaneous electric polarization promotes the desirable separation of photo-excited carriers and allows voltages higher than the bandgap, which may enable efficiencies beyond the maximum possible in a conventional p–n junction solar cell[2,6,8–10]. Ferroelectric oxides are also stable in a wide range of mechanical, chemical and thermal conditions and can be fabricated using low-cost methods such as sol–gel thin-film deposition and sputtering[3,5]. Recent work[3,5,11] has shown how a decrease in ferroelectric layer thickness and judicious engineering of domain structures and ferroelectric–electrode interfaces can greatly increase the current harvested from ferroelectric absorber materials, increasing the power conversion efficiency from about 10^{-4} to about 0.5 per cent. Further improvements in photovoltaic efficiency have been inhibited by the wide bandgaps (2.7–4 electronvolts) of ferroelectric oxides, which allow the use of only 8–20 per cent of the solar spectrum. Here we describe a family of single-phase solid oxide solutions made from low-cost and non-toxic elements using conventional solid-state methods: $[KNbO_3]_{1-x}[BaNi_{1/2}Nb_{1/2}O_{3-\delta}]_x$ (KBNNO). These oxides exhibit both ferroelectricity and a wide variation of direct bandgaps in the range 1.1–3.8 electronvolts. In particular, the $x = 0.1$ composition is polar at room temperature, has a direct bandgap of 1.39 electronvolts and has a photocurrent density approximately 50 times larger than that of the classic ferroelectric $(Pb,La)(Zr,Ti)O_3$ material. The ability of KBNNO to absorb three to six times more solar energy than the current ferroelectric materials suggests a route to viable ferroelectric semiconductor-based cells for solar energy conversion and other applications.

The wide bandgap of typical ferroelectric perovskites (with ABO_3 composition) is due to the fundamental characteristics of the metal–oxygen A–O and B–O bonds. The excitation across the bandgap is essentially a charge transfer from the oxygen (O) $2p$ states at the valence band maximum to the transition-metal d states at the conduction band minimum. Transition-metal B cations enable the perovskite oxide to exhibit ferroelectricity[12]. Owing to a large difference in electronegativity between the oxygen and transition-metal atoms, the bandgap is quite large (3–5 eV). The lowest known bandgap for a ferroelectric oxide has been $E_g = 2.7$ eV, obtained for $BiFeO_3$ and the recently fabricated $LaCoO_3$-doped $Bi_4Ti_3O_{12}$ films[6,13]. This made $BiFeO_3$ the subject of a number of investigations for photovoltaic applications[1,4,14]. However, $BiFeO_3$ is capable of absorbing only 20% of the solar spectrum, necessitating the development of new semiconducting ferroelectric oxides. For example, a weakly ferroelectric non-perovskite $KBiFe_2O_5$ material has recently been discovered with a bandgap of 1.6 eV (ref. 15).

Following the bandgap-engineering strategy explored in a previous theoretical study[16–18] on Ni-doped $PbTiO_3$, we used two different transition-metal cations on the perovskite B-site to create ferroelectric perovskites with low bandgaps, with one cation driving ferroelectricity and the other giving an E_g in the visible range. We used the classic ferroelectric perovskite $KNbO_3$ (KNO) to provide off-centre distortions and polarization ($P \approx 0.55\,\text{C}\,\text{m}^{-2}$ at 0 K)[19], and mix it with $BaNi_{1/2}Nb_{1/2}O_{3-\delta}$ (BNNO) to introduce a combination of Ni^{2+} on the B-site and an oxygen vacancy, which can give rise to electronic states in the gap of the parent KNO material. Nb-containing ferroelectric perovskites have been shown to tolerate a high concentration of vacancies[20] so B-site Nb ions should be able to accommodate the Ni^{2+}-oxygen vacancy combination. The large sizes of K and Ba cations favour solubility and vacancy formation, because Ni^{3+} has a small ionic radius and is only stable in perovskites with (small) La^{3+} cations on the A-site, whereas the larger Ni^{2+} ion is known to be stable in ferroelectric compounds such as $PbNi_{1/3}Nb_{2/3}O_3$ (ref. 21).

The solid solutions $[KNbO_3]_{1-x}[BaNi_{1/2}Nb_{1/2}O_{3-\delta}]_x$ with compositions $x = 0.1$–0.5 were synthesized by standard solid-state synthesis methods. The samples were sintered to 95% density and characterized. Synchrotron X-ray diffraction shows (Fig. 1b) the formation of a stable perovskite for all solutions, with very small NiO impurity peaks. The increase in lattice parameters with BNNO substitution is consistent with the presence of the Ni^{2+} cation, with a larger ionic radius (0.69 Å) than Nb^{5+} (0.64 Å) or Ni^{3+} (0.60 Å).

To examine the microscopic structure and properties of KBNNO, we performed first-principles density functional theory (DFT) calculations for the $x = 0.33$ composition using a 60-atom supercell (Fig. 1c). Two of the twelve Nb^{5+} ions are replaced by Ni^{2+}, and four of the twelve K ions are replaced by Ba. This substitution will generate an oxygen vacancy $V_O^{\bullet\bullet}$ adjacent to Ni_{Nb}''' defects (in Kröger–Vink notation)[22] with the local dipole (Ni–V_O) parallel to the overall polarization P. We obtained two stable KBNNO configurations, with the local structure of Ni^{2+}–V_O–Ni^{2+} and Ni^{2+}–V_O–Nb^{5+} (Fig. 1c). The calculated P values are $0.19\,\text{C}\,\text{m}^{-2}$ and $0.18\,\text{C}\,\text{m}^{-2}$ for the two KBNNO structures, mainly owing to the Nb off-centre distortions. The polarization is smaller than that of the parent KNO material ($P = 0.43\,\text{C}\,\text{m}^{-2}$ at 0 K in DFT) but still substantial. Comparison of lattice parameters for fully oxidized ($\delta = 0$) compositions (KBNNO$_\text{ox}$) and KBNNO ($\delta > 0$, with oxygen vacancies) showed that KBNNO volume is increased compared to KNO, in agreement with experimental data, whereas the KBNNO$_\text{ox}$ volume is decreased (Extended Data Table 1). At the synthesis conditions, DFT + U (where U is the Hubbard on-site repulsion term) free-energy calculations[23] found oxygen vacancies to be thermodynamically favoured, indicating that the material is in the KBNNO state. Finally, KBNNO samples are not conductive, in contrast to KBNNO$_\text{ox}$, for which the DFT + U calculations predict a metallic state. Therefore, the material we synthesized is indeed KBNNO, with oxygen vacancies.

For efficient and practical separation of excited carriers, ferroelectrics must be polar at room temperature and higher. Comparison of the 0-K, DFT-calculated P values for KNO and $x = 0.33$ KBNNO ($0.43\,\text{C}\,\text{m}^{-2}$ and about $0.2\,\text{C}\,\text{m}^{-2}$, respectively) shows that BNNO substitution decreases polarization. This should reduce the temperature

[1]The Makineni Theoretical Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, USA. [2]Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6272, USA. [3]Department of Materials Science and Engineering, Drexel University, Philadelphia, Pennsylvania 19104, USA.
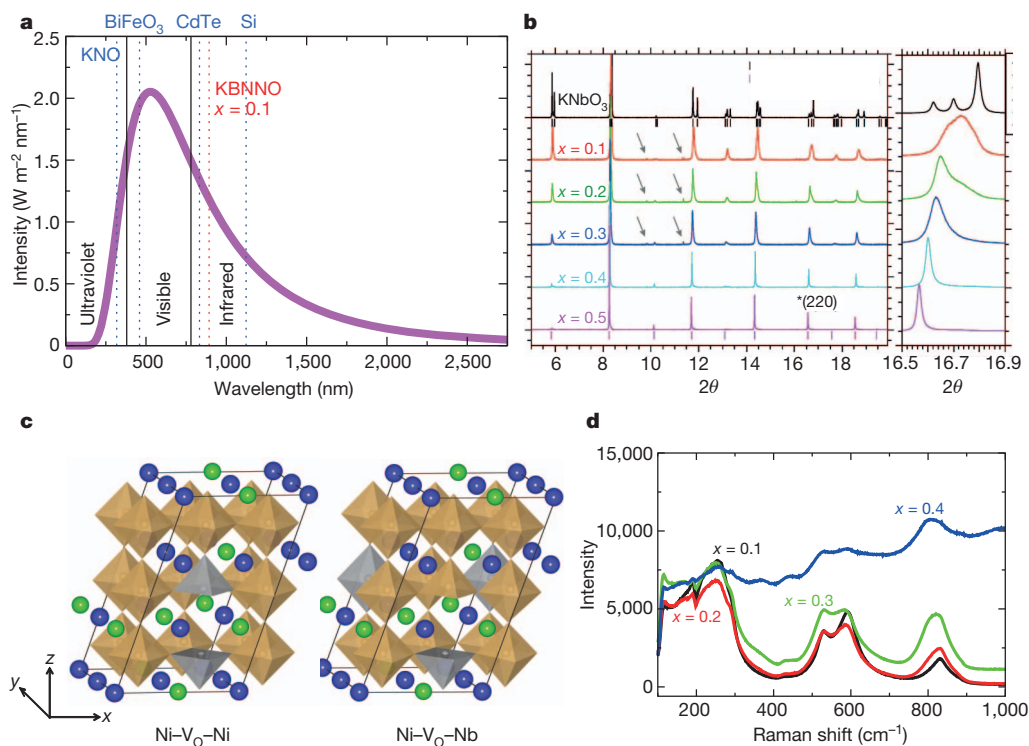
Figure 1 | KBNNO structural properties. a, The solar spectrum and E_g values for Si, CdTe, BiFeO$_3$ and $x = 0.1$ KBNNO. **b**, KBNNO synchrotron X-ray diffraction results. The inset shows the (220)p family of reflections. Perovskite peaks are marked by tick marks and NiO impurity peaks by arrows. **c**, (KNbO$_3$)$_8$–(BaNb$_{1/2}$Ni$_{1/2}$O$_{2.75}$)$_4$ crystal structures used in DFT calculations.

K and Ba are shown by blue and green spheres, respectively; Nb–O$_6$ and Ni–O$_6$ are shown as brown and grey octahedra with O atoms at the vertices. **d**, Raman data for KBNNO $x = 0.1$–0.4 compositions. For $x \leq 0.3$, a depth resonance at 200 cm^{-1} and a peak near 800 cm^{-1} indicate a ferroelectric phase[25].

of the ferroelectric-to-paraelectric and orthorhombic-to-tetragonal transitions and make the tetragonal phase preferred for ferroelectric KBNNO compositions at room temperature[24]. X-ray-diffraction data for the (220)p (where 'p' represents the cubic perovskite sub-cell) peak for $x = 0$–0.5 compositions shows a gradual transition from the orthorhombic ferroelectric KNO structure to a cubic structure at $x = 0.4$. For $x = 0.1$, the broad (220)p peak exhibits a shoulder on the side of lower 2θ (scattering angle); this is consistent with a weakly tetragonal ferroelectric phase. Raman spectroscopy (Fig. 1d) shows a resonance depth at 200 cm^{-1} and a peak at 820 cm^{-1} for $x \leq 0.3$ compositions; these have been identified as signatures of ferroelectricity in KNbO$_3$-based solid solutions[25]. These features are sharpest for the $x = 0.1$ composition, which also exhibits peaks in the dielectric constant at around 450 K and at around 600 K (Extended Data Fig. 1a) and is the focus of our further investigation.

The ferroelectric switching measurements on a 20-μm-thick $x = 0.1$ sample in high vacuum (pressure of 10^{-7} torr) at 77–170 K showed standard ferroelectric hysteresis loops reaching the maximum value of about 0.2 C m^{-2} at 170 K (Fig. 2a, Extended Fig. 1b) for an applied field of 250 kV cm^{-1}. We ascribe the increase of measured P with temperature to the greater speed of domain-wall motion and, therefore, more effective switching at higher temperatures. The measured P value is therefore the lower limit of the true bulk P. With further increase in temperature, increased leakage made poling ineffective. Under ambient conditions, poling of the sample produced very thin, elongated loops with P reaching only 0.01 C m^{-2} (Extended Data Fig. 1c, d). In contrast to the hysteresis measurements at room temperature (300 K), local ferroelectric piezoelectric measurements on a thin and electrically addressable lamella extracted (see Methods) from an $x = 0.1$ sample (Fig. 2b) showed a strong switching loop characteristic of ferroelectric materials, as has been found before for leaky ferroelectrics[1,15]. Taken together, the DFT calculations and experimental data unambiguously show the presence of strong P in KBNNO for $x = 0.1$.

We characterize the light-absorption properties of the KBNNO pellets using spectroscopic ellipsometry (Fig. 2c). We find that the bandgaps of Ni-containing KBNNO solutions are in the range 1.1–2.0 eV; this is much lower than the 3.8-eV bandgap of the KNO material (Fig. 2d). Owing to the smaller bandgap, the samples are green, in contrast to KNO, which is white. The bandgap tunability of 2.7 eV is 150% greater than previously achieved by doping of ferroelectric Bi$_4$Ti$_3$O$_{12}$ or by the doping of the non-ferroelectric perovskite Ba(In$_{1/2}$Ta$_{1/2}$)O$_3$ and is on par with the largest bandgap tunabilities observed in oxides (such as $E_g = 1.4$–3.9 eV variation in CdO–CaO solid solutions)[6,26–28]. To our knowledge, the $E_g = 1.1$–3.8 eV variation of KBNNO solid solutions is the largest ever observed for a perovskite or a ferroelectric material. The bandgaps are direct, as indicated by a single slope of the extinction coefficient versus wavelength, and the power law of its variation. The absorption coefficient is approximately 2.5×10^4 cm^{-1} at 885 nm, comparable to the absorption coefficient of CdTe and GaAs. Inspection of Fig. 2d shows that there is a non-monotonic change in E_g with BNNO fraction, with an initial steep decay for low BNNO fraction and then a slow rise starting from $x = 0.3$.

To elucidate the origin of the bandgap lowering in the KBNNO solid solutions, we examined the electronic structure of KBNNO with first-principles methods. The electronic structures of the $x = 0.33$ KBNNO supercells show direct bandgaps of 1.84 eV and 1.49 eV, much smaller than the 2.3-eV local density approximation (LDA) + U bandgap of KNO (Extended Data Fig. 2). The valence band maximum consists of hybridized Ni $3d$ and O $2p$ states, while the conduction band minimum is composed of Nb $4d$ states. The filled Ni $3d$ gap states in KBNNO therefore play a crucial part in lowering the bandgap.

Photoresponse measurements showed that KBNNO is promising for photovoltaic applications. We first examined the dependence of the photocurrent response of KBNNO on incident optical wavelength, using a monochromatic source tunable from 700 to 900 nm (Fig. 3a).
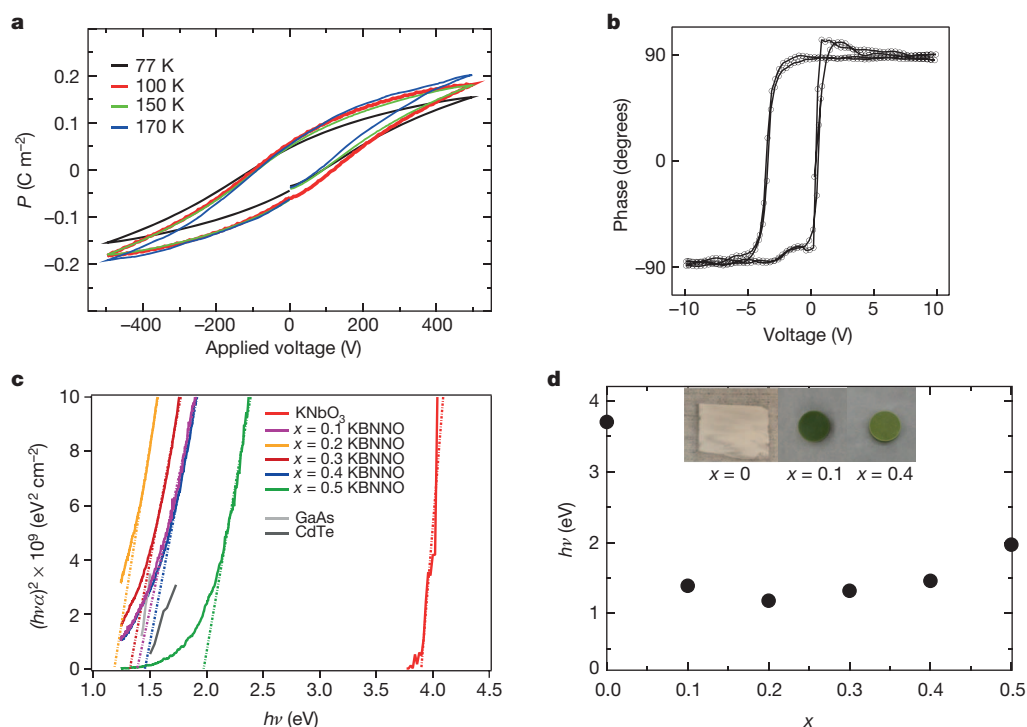
Figure 2 | Experimental results for ferroelectricity and bandgap in KBNNO. a, Ferroelectric hysteresis loops for a 20-μm-thick $x = 0.1$ KBNNO film at 10^{-7} torr and 77–170 K. **b**, Local ferroelectric piezoelectric hysteresis loops measured for an $x = 0.1$ lamella sample of KBNNO. **c**, Ellipsometry measurements for KBNNO oxides with $x = 0.0$–0.5, showing bandgaps from 1.18 eV to 3.8 eV. This makes KBNNO promising for visible solar light absorption. **d**, Bandgap values versus BNNO fraction. Also shown are images of the KNbO$_3$ and KBNNO pellets for $x = 0.1$ and $x = 0.4$ compositions.

The response starts to rise at about 850 nm (1.46 eV) and peaks at about 710 nm (1.74 eV), showing a good match with the solar spectrum. Measurements of the open-circuit photovoltage V_{oc} and short-circuit

photocurrent J_{sc} found that the photoresponse of KBNNO is controlled by the polarization and is much larger than that for the classic ferro-electric (Pb,La)(Zr,Ti)O$_3$ and (Na,K)NbO$_3$ materials. The 20-μm-thick ceramic sample was first poled at 77 K with a 500 V pulse for 400 s, and its response was then measured in the dark and under illumination by a halogen lamp delivering about 4 mW cm^{-2} of above-bandgap illumination (Fig. 3b). The direction of the photocurrent is reversed after the material is poled in the opposite direction; this is a signature of excited carrier separation by the bulk of material exhibited by ferroelectrics. The measured J_{sc} and V_{oc} are about 40 nA cm^{-2} and 3.5 V, respectively. The large value of V_{oc} is in line with previous reports of above-bandgap photovoltage in ferroelectric materials[29].

We then measured V_{oc} and J_{sc} at room temperature after poling with a ±80-V pulse for 300 s. We note that, on the basis of the hysteresis-loop measurements discussed above, this procedure will pole only a small fraction of the sample. We found V_{oc} and J_{sc} of 0.7 mV and about 0.1 μA cm^{-2}, respectively. Here too, we found a reversal of photocurrent direction on change in the sign of the poling voltage (Extended Data Fig. 3). Despite the weak poling at room temperature, the 300-K J_{sc} is greater than the 77-K J_{sc}. This is due to the strong dependence of J_{sc} on temperature[29]. The room-temperature J_{sc} is higher than 8 nA cm^{-2} for a 50-μm (Pb,La)(Zr,Ti)O$_3$ sample or 25 nA cm^{-2} for 0.84-μm (Na,K)NbO$_3$ samples measured in previous 300-K experiments under ultraviolet illumination[30,31]. Our 300-K KBNNO results also compare favourably to the photoresponse of BiFeO$_3$ reported[1] for a 70-μm sample under green-light illumination ($J_{sc} = 4$ μA cm^{-2}, $V_{oc} = 35$ mV for 10 mW cm^{-2} illumination) considering the broad-spectrum illumination and the partial poling for our KBNNO sample.
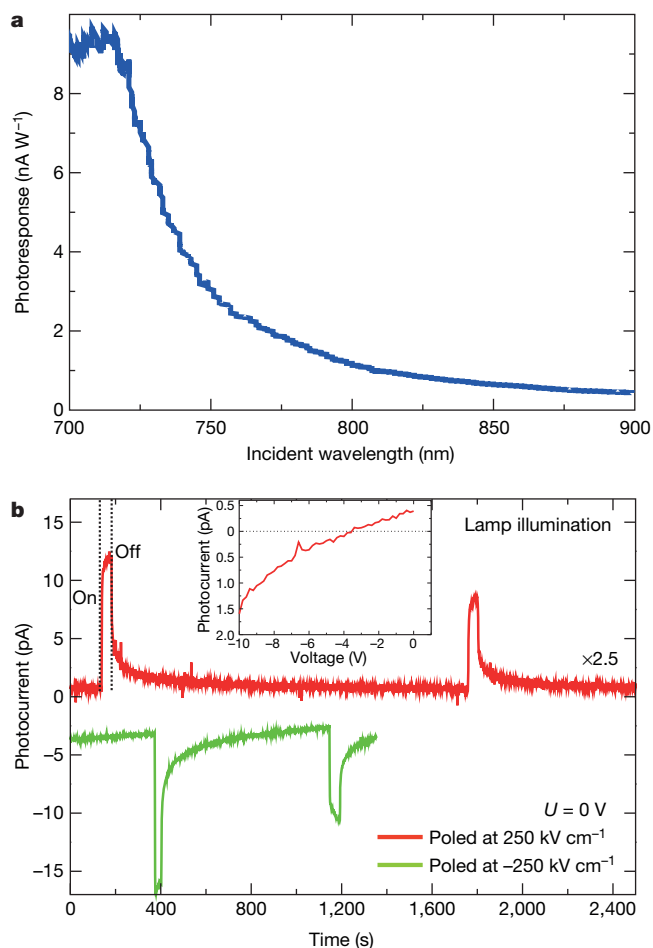
Figure 3 | Photocurrent measurements for the KBNNO samples. a, The current collected between two co-planar 85-μm^2 electrodes per watt of total incident illumination. The photoresponse starts at the bandgap energy of 1.39 eV and saturates at 1.74 eV. **b**, Ferroelectric photovoltaic effect under short circuit ($U = 0$) conditions for 20-μm-thick $x = 0.1$ film at 77 K following poling by a ±500-V pulse applied for 400 s and under 4 mW cm^{-2} of above-bandgap illumination. Reversal of poling voltage results in the reversal of photocurrent direction. The inset shows the photoresponse versus applied bias at 77 K obtained by subtracting light current from dark; a large V_{oc} of 3.5 V is observed.

Decreased thickness of the ferroelectric layer and optimization of the ferroelectric–metal interfaces have been shown to increase the photocurrent of wide-bandgap ferroelectric-based solar cells by up to six orders of magnitude[3,5]. In particular, 270-nm-thick PZT-based cells with a Cu_2O cathode buffer layer have been demonstrated to reach a tenth of the theoretically possible efficiency (0.57% for the case of PZT with $E_g = 3.5$ eV.) The ideal match of the KBNNO bandgap to the solar spectrum, its compositional tuning throughout the visible range and its photoresponse properties open up the possibility of ferroelectric photovoltaic efficiency of >3% in a thin-film device and the use of ferroelectric materials as solar absorber layers and carrier separators in practical photovoltaics. It is also important for the emerging field of ferroelectric photovoltaics as the first visible-light-absorbing strongly ferroelectric material.

METHODS SUMMARY

All samples were made by standard solid-state synthesis techniques in the powder form, followed by sintering. An integrated focused ion beam and scanning electron microscope (FEI, DB235) equipped with a lift-out tool (Omniprobe) was used to extract thin-film lamellae from the bulk-synthesized KBNNO pellets for bandgap and local ferroelectric piezoelectric switching measurements. Mesoscopic (about 20 μm thick) samples were prepared by metallization of the polished side of each pellet with thermally evaporated layers of Cr (5 nm) and Au (100 nm), followed by mounting with the pellet's metallized side face down, thinning and subsequent polishing to a root mean squared roughness of around 100 nm. Indium/tin oxide thin films (about 50 nm) were subsequently deposited onto shadow-masked, approximately 20-μm-thick KBNNO via pulsed laser deposition at 200 °C in O_2. Bandgaps were measured by spectroscopic ellipsometry using a variable-angle spectroscopic ellipsometer equipped with Glan–Taylor polarizers, a rotating compensator, and deuterium and quartz halogen lamps for spectral coverage (J. A. Woollam, Model M2000). Ferroelectric switching within the lamellae was evaluated from the local piezoresponse using a scanning probe microscope (Asylum Research, MFP-3D) and Pt-coated Si probes (Olympus, AC 240TM; nominal stiffness constant, about $2\,N\,m^{-1}$). Ferroelectric polarization hysteresis measurements were collected at 77–200 K in high vacuum (10^{-7} torr) and at ambient pressure using a ferroelectric tester (Radiant LC). Steady-state photocurrent/bias-voltage traces were collected in bulk KBNNO and BTO and in approximately 20-μm-thick films under 120-W tungsten–halogen spectrally broad lamp probe illumination and using a fibre-coupled supercontinuum source (NKT Compact), under vacuum (10^{-6} torr, Lakeshore Cryotronics, TTP4) using a picoammeter (Keithley, model 6487). Photocurrent spectra were collected using a tunable-wavelength Ti:sapphire laser (M2 SolsTiS) with an incident spot diameter of about 10 mm. First-principles DFT LDA + U calculations were done using norm-conserving pseudopotentials and a plane-wave basis set, as implemented in the Quantum-Espresso package.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Choi, T., Lee, S., Choi, Y., Kiryukhin, V. & Cheong, S.-W. Switchable ferroelectric diode and photovoltaic effect in $BiFeO_3$. *Science* **324,** 63–66 (2009).
2. Yang, S. Y. *et al.* Above-bandgap voltages from ferroelectric photovoltaic devices. *Nature Nanotechnol.* **5,** 143–147 (2010).
3. Cao, D. *et al.* High-efficiency ferroelectric-film solar cells with an n-type Cu_2O cathode buffer layer. *Nano Lett.* **12,** 2803–2809 (2012).
4. Alexe, M. & Hesse, D. Tip-enhanced photovoltaic effects in bismuth ferrite. *Nature Commun.* **2,** 256 (2011).
5. Qin, M., Ao, K. & Liang, Y. C. High efficiency photovoltaics in nanoscaled ferroelectric thin films. *Appl. Phys. Lett.* **93,** 122904 (2008).
6. Choi, W. S. *et al.* Wide bandgap tunability in complex transition metal oxides by site-specific substitution. *Nature Commun.* **3,** 689 (2012).
7. Kreisel, J., Alexe, M. & Thomas, P. A. A photoferroelectric material is more than the sum of its parts. *Nature Mater.* **11,** 260 (2012).
8. Fridkin, V. M. *Photoferroelectrics* (Springer, 1979).
9. Inoue, Y., Sato, K., Sato, K. & Miyama, H. Photoassisted water decomposition by ferroelectric lead zirconate titanate ceramics with anomalous photovoltaic effects. *J. Phys. Chem.* **90,** 2809–2810 (1986).
10. Young, S. M. & Rappe, A. M. First principles calculation of the shift current photovoltaic effect in ferroelectrics. *Phys. Rev. Lett.* **109,** 116601 (2012).
11. Glass, A. M., Linde, D. V. D. & Negran, T. J. High-voltage bulk photovoltaic effect and photorefractive process in $LiNbO_3$. *Appl. Phys. Lett.* **25,** 233–235 (1974).
12. Cohen, R. E. Origin of ferroelectricity in perovskite oxides. *Nature* **358,** 136–138 (1992).
13. Basu, S. R. *et al.* Photoconductivity in $BiFeO_3$ thin films. *Appl. Phys. Lett.* **92,** 091905 (2008).
14. Yang, S. Y. *et al.* Photovoltaic effects in $BiFeO_3$. *Appl. Phys. Lett.* **95,** 062909 (2009).
15. Zhang, G. *et al.* New high T_c multiferroics $KBiFe_2O_5$ with narrow band gap and promising photovoltaic effect. *Sci. Rep.* **3,** 1265 (2013).
16. Bennett, J. W., Grinberg, I. & Rappe, A. M. New highly polar semiconductor ferroelectrics through d^8 cation-O vacancy substitution into $PbTiO_3$: a theoretical study. *J. Am. Chem. Soc.* **130,** 17409–17412 (2008).
17. Gou, G. Y., Bennett, J. W., Takenaka, H. & Rappe, A. M. Post density functional theoretical studies of highly polar semiconductive $Pb(Ti_{1-x}Ni_x)O_{3-x}$ solid solutions: effects of cation arrangement on band gap. *Phys. Rev. B* **83,** 205115 (2011).
18. Qi, T., Grinberg, I. & Rappe, A. M. Band-gap engineering via local environment in complex oxides. *Phys. Rev. B* **83,** 224108 (2011).
19. Liang, L., Li, Y. L., Chen, L.-Q., Hu, S. Y. & Lu, G.-H. A thermodynamic free energy function for potassium niobate. *Appl. Phys. Lett.* **94,** 072904 (2009).
20. Zhao, X., Qu, W. & Tan, X. Zr-modified $Pb(Mg_{1/3}Nb_{2/3})O_3$ with a long-range cation order. *J. Am. Ceram. Soc.* **91,** 3031–3038 (2008).
21. Kondo, M. & Kurihara, K. Sintering behavior and surface microstructure of PbO-rich $Pb(Ni_{1/3}Nb_{2/3})O_3–PbZrO_3$ ceramics. *J. Am. Ceram. Soc.* **84,** 2469–2474 (2001).
22. Kröger, F. A. & Vink, H. J. Relations between the concentrations of imperfections in crystalline solids. **3,** 307–435 (1956).
23. Reuter, K. & Scheffler, M. First-principles atomistic thermodynamics for oxidation catalysis: surface phase diagrams and catalytically interesting regions. *Phys. Rev. Lett.* **90,** 046103 (2003).
24. Grinberg, I. & Rappe, A. M. Local structure and macroscopic properties in $Pb(Zn_{1/3}Nb_{2/3})O_3–PbTiO_3$ and $Pb(Mg_{1/3}Nb_{2/3})O_3–PbTiO_3$ solid solutions. *Phys. Rev. B* **70,** 220101 (2004).
25. Bartasyte, A., Kreisel, J., Peng, W. & Guilloux-Viry, M. Temperature-dependent Raman scattering of $KTa_{1-x}Nb_xO_3$ thin films. *Appl. Phys. Lett.* **96,** 262903 (2010).
26. Srihari, V. *et al.* Wide band gap tunability of bulk $Cd_{1-x}Ca_xO$. *J. Appl. Phys.* **109,** 013510 (2011).
27. Ji, S. M., Choi, S. H., Jang, J. S., Kim, E. S. & Lee, J. S. Band gap tailored $Zn(Nb_{1-x}V_x)O_6$ solid solutions as visible light photocatalysts. *J. Phys. Chem. C* **113,** 17824–17830 (2009).
28. Kim, T. W. *et al.* Substitution effect of pentavalent bismuth ions on electronic structure and physicochemical properties of perovskite-structured $Ba(In_{0.5}Ta_{0.5})O_3$ semiconductors. *Mater. Res. Bull.* **42,** 1914–1920 (2007).
29. Brody, P. S. Temperature dependence of the short circuit photocurrent in ferroelectric ceramics. *Ferroelectrics* **10,** 143–146 (1976).
30. Poosanaas, P., Dogan, A., Thakoor, S. & Uchino, K. Influence of sample thickness on the performance of photostrictive ceramics. *J. Appl. Phys.* **84,** 1508–1512 (1998).
31. Park, J., Won, S. S., Ahn, C. W. & Kim, I. W. Ferroelectric photocurrent effect in polycrystalline lead-free $(K_{0.5}Na_{0.5})(Mn_{0.005}Nb_{0.995})O_3$ thin film. *J. Am. Ceram. Soc.* **96,** 146–150 (2013).

Author Contributions I.G. and A.M.R. created the materials design strategy. D.V.W. and P.K.D. suggested the KBNNO composition. I.G., J.E.S., P.K.D. and A.M.R. designed the calculations and experiments and supervised the analysis of obtained results. D.V.W., D.M.S. and L.W. synthesized the KBNNO powders and pellets. D.V.W. obtained the X-ray diffraction and dielectric data. G.C. developed the procedure to prepare the lamellae from the pellets. M.T. performed the piezoresponse and ellipsometry measurements. A.R.A. and J.E.S. analysed the Raman spectra. A.R.A., G.C., E.M.G. and J.E.S. carried out the ferroelectric, photoresponse and photovoltage measurements. G.G. performed the DFT calculations. I.G., G.G., J.E.S., A.M.R. and P.K.D. co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.R. (rappe@sas.upenn.edu).

METHODS

Synthesis and dielectric measurements. All samples were made from stoichiometric quantities of dried K_2CO_3, $BaCO_3$, NiO and Nb_2O_5 powders. After mixing in a mortar, the powders were ball-milled using yttria-stabilized zirconia planetary milling media in ethanol for 2 h. The dried powders were calcined on Pt foil in an alumina crucible at 900 °C for 12 h. Approximately 300 mg aliquots were pressed into 0.25-ml pellets in a uniaxial press and isostatically pressed at 80,000 psi (pounds per square inch). The pellets were placed on Pt foil in a covered alumina crucible, surrounded by sacrificial powder of the same composition to inhibit volatilization of potassium, and sintered at temperatures between 1,050 °C and 1,250 °C, depending on the composition. To minimize any absorption of H_2O, which is a potential issue in the synthesis of $KNbO_3$, at all stages of the synthesis samples were kept either at elevated temperature (at least 200 °C) or in a desiccator to minimize their exposure to moisture. Powder X-ray-diffraction patterns of the samples were collected on a laboratory X-ray diffractometer (Rigaku GiegerFlex D/Max-B) using Cu Kα radiation generated at 45 kV and 30 mA and by synchrotron X-ray diffraction (wavelength 0.413473 Å) using the Advanced Photon Source at Argonne National Laboratory. The dielectric data were collected on pellets coated with Ag paint (Heraeus ST1601-14 type) to provide electrical contacts for the Pt lead wires. The dielectric properties were investigated as functions of frequency and temperature using a high-precision impedance–capacitance–resistance meter (Hewlett-Packard, model 4284A) and a high-temperature thermal chamber. The sample temperature was monitored by an S-type thermocouple positioned near the pellet.

Spectroscopic ellipsometry. Spectroscopic ellipsometry was performed on polished KBNNO at 300 K in the 247–1,000-nm wavelength range using a variable-angle spectroscopic ellipsometer equipped with Glan–Taylor polarizers, a rotating compensator, and deuterium and quartz halogen lamps for spectral coverage (J. A. Woollam, model M2000). Measurement of the components of linearly polarized reflectivity at each selected wavelength were used to obtain the ellipsometric parameters Ψ and Δ through the relation

$$\tan\Psi(\lambda)\exp(i\Delta(\lambda)) = R_p(\lambda)/R_s(\lambda)$$

where $R_p(\lambda)$ and $R_s(\lambda)$ are reflection coefficients for light polarization parallel and perpendicular to the plane of incidence, respectively. The energy-dependent complex dielectric function was calculated using Fresnel's equations. The bandgap was calculated using a Tauc plot of $(h\nu\alpha)^2$ versus $h\nu$, where α is the absorption coefficient. Measurements were taken at 65°.

Extraction of thin-film lamellae. An integrated focused ion beam and scanning electron microscope (FEI, DB235) equipped with a lift-out tool (Omniprobe) was used to extract thin-film lamellae from the bulk-synthesized KBNNO. Briefly, a thin layer of carbon by sputtering coating (several tens of nanometres) is first deposited to provide protection against subsequent ion-beam irradiation and to enhance the imaging contrast. This was followed by deposition of a 500-nm-thick platinum film using ion-beam-assisted deposition onto the lift-out area, preventing direct ion-beam damage during the process. The lift-out preparation process consists of initial cross-sectional milling steps on both sides, a series of thinning steps using lower ion-beam currents, a finer-scale cross-sectional cleaning using an approximately 100-pA ion-beam current, and ion-beam local deposition of Pt to affix the lamella to the lift-out tool. A low beam current (<100 pA) was maintained during the final release of the lamella from the substrate. Using the lift-out probe, each harvested lamella was transferred carefully to glass substrates coated with layers of fluorine-doped tin oxide (TEC-15, Pilkington) and a top coating of 30 nm of indium (selected to facilitate wetting of the bottom contact to the lamella) deposited via electron-beam evaporation in vacuum. Each lamella is transferred to the substrate with the lamella first making contact with the substrate along one edge; the free-standing lamella is then pushed down onto the surface. As a final step in the transfer of the lamellar thin-film test specimen, Pt is deposited using electron-beam-assisted deposition onto the corners of each lamella to anchor it. Following this, post-processing steps of ultralow-beam-current surface ion milling and subsequent thermal annealing (500 °C for 5 h, followed by a slow cooling at 1 °C s^{-1}) were carried out in a furnace (Ney Vulcan 3-130) to effectively eliminate ion damage.

Local ferroelectric measurements. The KBNNO pellets were cut to roughly 250 μm with a diamond saw and polished under water to thicknesses of about 25 μm using lapping films (3M) coated with successively finer aluminium oxide particles. The final polish was done using a slurry of 0.05-μm colloidal silica (Ted Pella) in an alkaline suspension (pH 9.8). We estimate that the surface roughness should be less than 0.1 μm.

Ferroelectric switching within the lamellae was evaluated from the local piezoresponse using a scanning probe microscope (Asylum Research MFP-3D) and Pt-coated Si probes (AC 240TM, Olympus; nominal stiffness constant, about 2 N m^{-1}). A triangular waveform (frequency 0.025 Hz, peak-to-peak bias of 10 V) was applied to the bottom electrode while a sinusoidal alternating-current probing voltage (5 kHz, 0.5 V amplitude) was applied to the cantilever tip to collect the variation in the cantilever phase as a function of the bias voltage. The cantilever phase signal at the modulation frequency was collected with the aid of a digital lock-in amplifier (Stanford Research Systems SR830).

Ferroelectric hysteresis in mesoscopic (about 20 μm thick) KBNNO film samples ($x = 0.1$) was carried out at 77–200 K under 10^{-7} torr and at 300 K under ambient pressure in a probe station (Lakeshore Desert Cryotronics TTP4) using a ferroelectric tester (model LC, Radiant Technologies) and a high-voltage amplifier with selected bias voltage sweep rate periods ranging from 10 ms to 10 s and selected peak voltages of up to 500 V, and poling using direct-current bias for different durations ranging from 5 s to 400 s.

Photocurrent and Raman measurements. Electrical contacts were produced on KBNNO samples and $BaTiO_3$ samples using a shadow mask and Cr/Au layers, and on mesoscopic (about 20 μm thick) and polished films using thermally evaporated 200 μm × 200 μm Cr–Au (bottom side) and ITO deposited by pulsed laser deposition at 200 °C using shadow masks. The resulting structures on bulk were 85 μm × 85 μm pads separated by 45-μm gaps. Steady-state photocurrent/bias-voltage traces were collected under a halogen spectrally broad lamp probe illuminator (Dolan-Jenner MI-150), and alternately using a broadband supercontinuum laser source (NKT Compact) under ambient pressure and vacuum (10^{-7} torr, Lakeshore Cryotronics Model TTP4) using a picoammeter (Keithley model 6487). Photocurrent spectra were collected over the range of 700 nm to 900 nm using a wavelength-tunable Ti:sapphire laser (M2 SolsTiS). The laser radiation incident on the sample was about 10 mm in diameter, resulting in an incident intensity of 120 mW cm^{-2}. The laser spot was directed on the sample and aligned by maximizing the resulting output current. The $BaTiO_3$ photocurrent measurements were carried out on (100)-oriented substrate-grade $BaTiO_3$ (MTI Corporation). Raman scattering was collected at 300 K using the 543.5-nm laser line (Horiba Jobin-Yvon). Raman spectra were collected at 300 K under ambient pressure from a $x = 0.1$–0.4 bulk KBNNO sample and from $KNbO_3$.

Computational modelling. We perform first-principles calculations with a plane-wave basis set, as implemented in Quantum-Espresso[32]. The LDA exchange-correlation functional is used for structural relaxations, with a $6 \times 6 \times 6$ Monkhorst–Pack k-point grid and a 50-Ry plane-wave cut-off. All atoms are represented by norm-conserving optimized nonlocal pseudopotentials, generated with the OPIUM code (http://opium.sourceforge.net). The electronic contribution to the polarization is calculated following the Berry's phase formalism.

Because LDA severely underestimates the bandgap, and even falsely predicts KBNNO to be metallic, all the electronic structure calculations have been done at the level of LDA + U. Although LDA + U is unable to predict E_g with the accuracy of the more advanced hybrid functionals or GW[33] methods, it can still provide a good description of the change of E_g with respect to the solid-solution cation ordering[17]. A simplified version of the rotationally invariant formulation of the LDA + U method is employed in the present work, where U can be determined by self-consistent linear-response calculations. Under the conditions of our synthesis, LDA + U free-energy calculations[23] show oxygen vacancies to be thermodynamically favoured.

The dependence of bandgap on composition is due to the interplay between local bonding and the bandgap in KBNNO, as elucidated by LDA + U calculations. There are two possible configurations for the oxygen vacancies in KBNNO, Ni–V_O–Ni and Ni–V_O–Nb. Our calculations show that although both Ni–V_O–Ni and Ni–V_O–Nb configurations result in a lower bandgap owing to the introduction of the Ni $3d$ states, an extra density-of-states peak, contributed by the d-orbitals of the six-fold-coordinated Ni, is present in the valence-band maximum in Ni–V_O–Nb (see Supplementary Fig. 3). Therefore the E_g of the Ni–V_O–Nb configuration is lower than that of the Ni–V_O–Ni configuration. This configuration is also found to be more energetically favourable by our calculations. At low BNNO concentration, the Ni cations are isolated, so the Ni–V_O–Nb arrangement is prevalent. As Ni concentration increases, more Ni–V_O–Ni configurations are formed and therefore the measured bandgap E_g rises.

32. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter* **21,** 395502 (2009).

33. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. *Phys. Rev. B* **34,** 5390–5413 (1986).
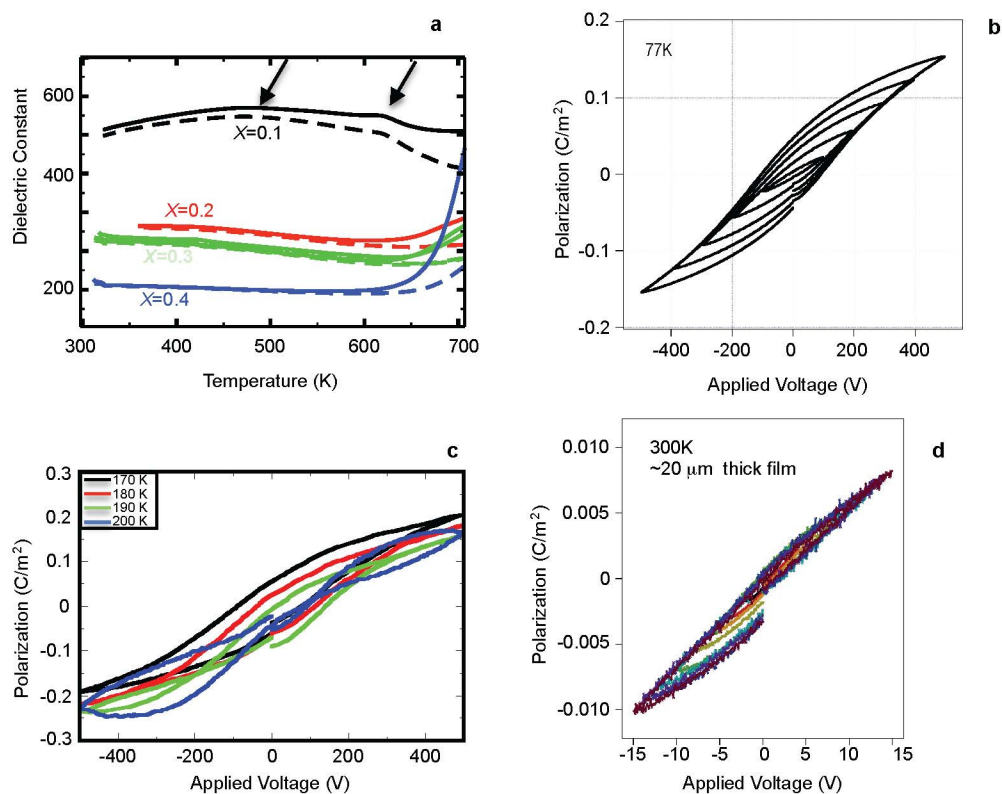
Extended Data Figure 1 | Ferroelectric and dielectric data. a, Dielectric data for $x = 0.1$–0.4 KBNNO. Two dielectric anomalies (arrows) at about 450 K and about 600 K are present (solid lines indicate heating; dotted lines indicate cooling). **b**, Ferroelectric hysteresis loops at 77 K, showing the effect of increasing the maximum poling voltage. **c**, Ferroelectric hysteresis loop for approximately 20-μm-thick $x = 0.1$ KBNNO film at 170–200 K. **d**, Ferroelectric hysteresis loop for approximately 20-μm-thick $x = 0.1$ KBNNO film at 300 K.
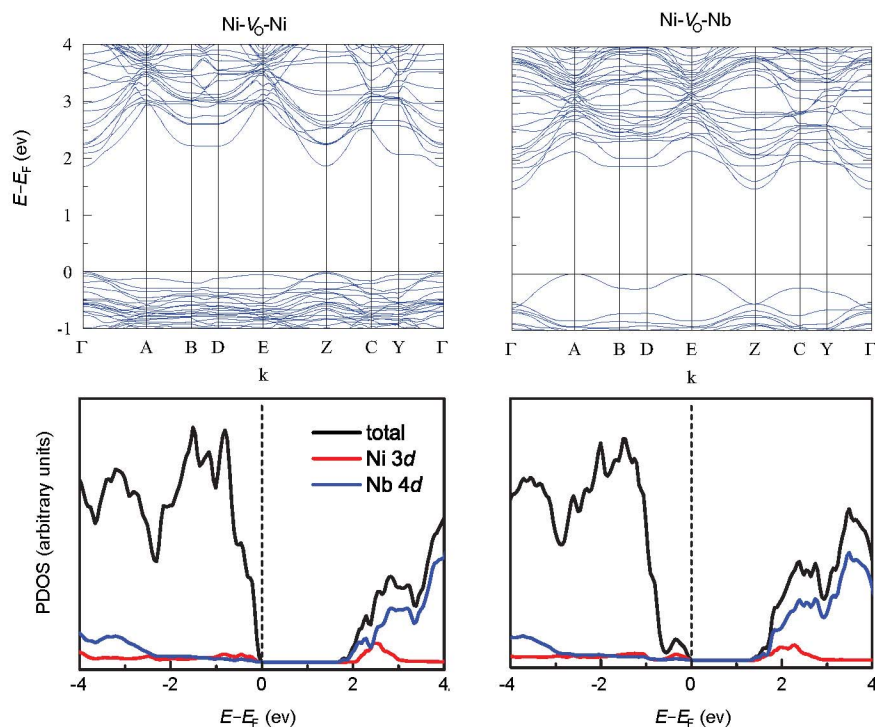
Extended Data Figure 2 | Electronic Structure of KBNNO. Band structures (top) and orbital-projected density of states (PDOS, bottom) for KBNNO Ni–V_O–Ni and Ni–V_O–Nb solid solutions near the Fermi level. The high-symmetry points in the Brillouin zone are Γ $(0, 0, 0)$ A $(-0.5, 0.5, 0)$, B $(-0.5, 0, 0)$, D $(-0.5, 0, 0.5)$, E $(-0.5, 0.5, 0.5)$, Z $(0, 0, 0.5)$, C $(0, 0.5, 0.5)$ and Y $(0, 0.5, 0)$. k is the wavevector. The more stable Ni–V_O–Nb structure provides a smaller bandgap. As Ni concentration rises, Ni–V_O–Ni becomes more common and the bandgap energy rises.
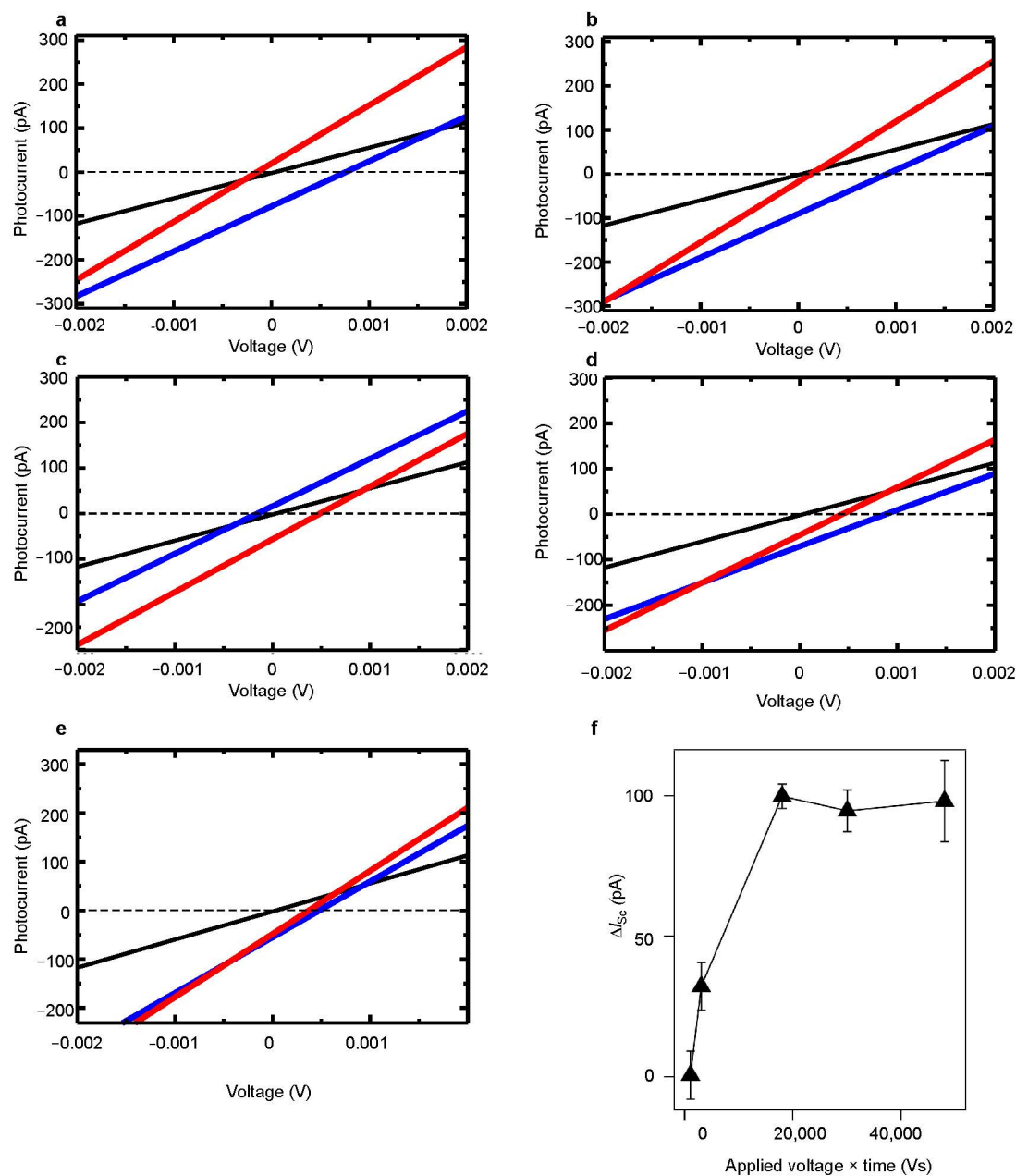
Extended Data Figure 3 | Switchable bulk photovoltaic effect in KBNNO and the dependence of photocurrent on poling. Ferroelectric photovoltaic effect for approximately 20-μm-thick $x = 0.1$ KBNNO film in ambient conditions under $4\,\mathrm{mW\,cm^{-2}}$ of above-bandgap illumination following poling by an 80-V pulse applied for 300 s (**a**), a 50-V pulse applied for 300 s (**b**), a 50-V pulse applied for 180 s (**c**), a 50-V pulse applied for 30 s (**d**) and a 50-V pulse applied for 10 s under $4\,\mathrm{mW\,cm^{-2}}$ of above-bandgap illumination (**e**). Black denotes collected dark current; blue and red traces indicate photocurrent following poling under positive and negative voltages, respectively. **f**, Short-circuit photocurrent I_{sc} for different product of duration and magnitude of poling voltage. The current is collected through 200 μm × 200 μm ITO and Cr–Au electrodes on the top and bottom of the sample, respectively. The height of each error bar is two standard deviations in the measured short-circuit current. As the applied voltage and poling time are increased, the difference between the photocurrents for the up- and down-polarized sample increases and the photocurrent magnitude rises by two orders of magnitude until saturation caused by leakage. This indicates that the sample is not yet fully poled even for the highest voltage possible in our set-up. Therefore, our results are the lower limit for the photocurrent for a fully poled material that can be achieved by application of larger electric fields in thinner films.

Extended Data Table 1 | Comparison of structural data from experiment and DFT calculations

		a (Å)	V (Å^3)
DFT	KNbO$_3$	3.985	63.26
	0.33 KBNNO Ni-Vac-Ni	3.986	63.35
	0.33 KBNNO Ni-Vac-Nb	3.99	63.64
	0.33 KBNNO Ni-O-Ni	3.975	62.86
	0.33 KBNNO Ni-O-Nb	3.971	62.59
Experiment	KNbO$_3$	4.015	64.70
	0.3 KBNNO	4.036	65.74

Pseudo-cubic lattice constant a and volume V values are as obtained computationally by DFT–LDA relaxations and by experimental X-ray diffraction measurements. DFT calculations for KBNNO with vacancies correctly reproduce the experimentally observed increase in the cell volume with increased BNNO content. In contrast, fully oxidized KBNNO samples show a decrease in cell volume compared to the parent KNbO$_3$ material. This indicates that the experimentally studied material is indeed KBNNO with vacancies.

LETTER

Origin and age of the earliest Martian crust from meteorite NWA 7533

M. Humayun[1], A. Nemchin[2]†, B. Zanda[3,4], R. H. Hewins[3,4], M. Grange[2], A. Kennedy[5], J.-P. Lorand[6], C. Göpel[7], C. Fieni[3], S. Pont[3] & D. Deldicque[8]

The ancient cratered terrain of the southern highlands of Mars is thought to hold clues to the planet's early differentiation[1,2], but until now no meteoritic regolith breccias have been recovered from Mars. Here we show that the meteorite Northwest Africa (NWA) 7533 (paired with meteorite NWA 7034[3]) is a polymict breccia consisting of a fine-grained interclast matrix containing clasts of igneous-textured rocks and fine-grained clast-laden impact melt rocks. High abundances of meteoritic siderophiles (for example nickel and iridium) found throughout the rock reach a level in the fine-grained portions equivalent to 5 per cent CI chondritic input, which is comparable to the highest levels found in lunar breccias. Furthermore, analyses of three leucocratic monzonite clasts show a correlation between nickel, iridium and magnesium consistent with differentiation from impact melts. Compositionally, all the fine-grained material is alkalic basalt, chemically identical (except for sulphur, chlorine and zinc) to soils from Gusev crater. Thus, we propose that NWA 7533 is a Martian regolith breccia. It contains zircons for which we measured an age of 4,428 ± 25 million years, which were later disturbed 1,712 ± 85 million years ago. This evidence for early crustal differentiation implies that the Martian crust, and its volatile inventory[4], formed in about the first 100 million years of Martian history, coeval with earliest crust formation on the Moon[5] and the Earth[6]. In addition, incompatible element abundances in clast-laden impact melt rocks and interclast matrix provide a geochemical estimate of the average thickness of the Martian crust (50 kilometres) comparable to that estimated geophysically[2,7].

NWA 7533 is a polymict breccia, characterized by a variety of clasts set in a fine-grained (~1 μm) interclast crystalline matrix (ICM) (Fig. 1). The main clast component consists of fine-grained (5–20 μm) clast-laden impact melt rock (CLIMR) occurring as oval or curved smooth bodies. Other clasts are made up of melt rock, melt spherules and fine-grained (20–100 μm) basaltic clasts, as well as lithic (noritic and mon-zonitic) and crystal (especially pyroxene and feldspar) clasts that occur in both melt rock and matrix (Fig. 1 and Supplementary Fig. 1). There is a surprising dearth of olivine in both matrix and clasts even though the Mg content of the matrix (~7.5%) is higher than that of Gusev crater soils. Among the lithic clasts are coarse-grained leucocratic rocks consisting of alkali feldspar, plagioclase, chlorapatite and ilmenite, with a monzonitic composition. Exsolution in both pyroxenes and alkali feldspars indicates that many lithic clasts are plutonic in origin (Supplementary Fig. 1). Chemical and oxygen isotopic[3] evidence confirms that NWA 7533 is a Martian meteorite (Supplementary Information). Here we present laser ablation ICP–MS (inductively coupled plasma mass spectrometry) elemental abundances and U–Pb zircon geochronology which demonstrate that NWA 7533 is a Martian regolith breccia, and discuss the implications of this result.

The ICM and CLIMR have abundances of Ni (400–700 p.p.m.) and Ir (10–80 p.p.b.) at their respective Mg contents (an index of chemical differentiation of basaltic liquids) that are much higher than those of shergottite–nakhlite–chassignite (SNC) meteorites (Ni < 200 p.p.m.; Ir < 1 p.p.b.) and comparable to those of lunar breccias[8,9] (Fig. 2), indicating a large meteoritic component. Moreover, the relative abundances of Ru, Rh and Os to Ir are in chondritic ratios in the ICM and CLIMR (Supplementary Fig. 2). The siderophile element contents of ICM and CLIMR require the equivalent of ~5% CI chondrite admixed into the Martian regolith. Prior explanations of the high Ni abundances in Gusev soils have included both indigenous[10] and meteoritic origins[11], but a chondritic impactor could not be inferred from Ni alone[11]. Surprisingly, the leucocratic clasts also have high Ni abundances relative to the SNC trend even at Mg < 0.1 wt%. The individual mineral spot analyses from two of the leucocratic clasts were examined after laser ablation ICP–MS analysis, and the spots were found to be contained entirely within the clast, not overlapping the Ni- and Ir-rich matrix (Supplementary Fig. 3). This is evidence that these leucocratic clasts
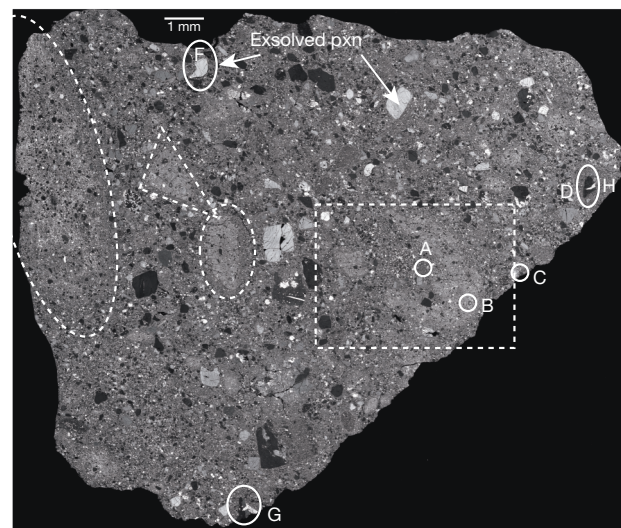
Figure 1 | Backscattered-electron image of NWA 7533 section 1. The breccia contains many large bodies of clast-laden impact melt rock (light or medium grey), some outlined with dot–dash lines, in fine-grained interclast crystalline matrix. Solid ellipses show crystal and lithic fragments, close-ups of which (lettered) are shown in Supplementary Information. Pyroxene (pxn; light or medium grey), feldspar (dark grey) and pyroxene–feldspar rock fragments are found in both melt rocks and matrix. Bright grey minerals include chlorapatite and Fe-rich oxides and oxyhydroxides.

[1]Department of Earth, Ocean and Atmospheric Science, and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310, USA. [2]Department of Applied Geology, Curtin University, Perth, Western Australia 6845, Australia. [3]Laboratoire de Minéralogie et Cosmochimie du Muséum, CNRS and Muséum National d'Histoire Naturelle, 75005 Paris, France. [4]Department of Earth and Planetary Sciences, Rutgers University, Piscataway, New Jersey 08854, USA. [5]Department of Applied Physics, Curtin University, Perth, Western Australia 6845, Australia. [6]Laboratoire de Planétologie et Géodynamique de Nantes, CNRS UMR 6112, Université de Nantes, 2 Rue de la Houssinière, BP 92208, 44322 Nantes Cedex 3, France. [7]Institut de Physique du Globe, Sorbonne Paris Cité, University Paris Diderot, CNRS UMR 7154, F-75005 Paris, France. [8]Ecole Normale Supérieure, UMR 8538, 75231 Paris Cedex 5, France. †Present address: Laboratory for Isotope Geology, Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden.
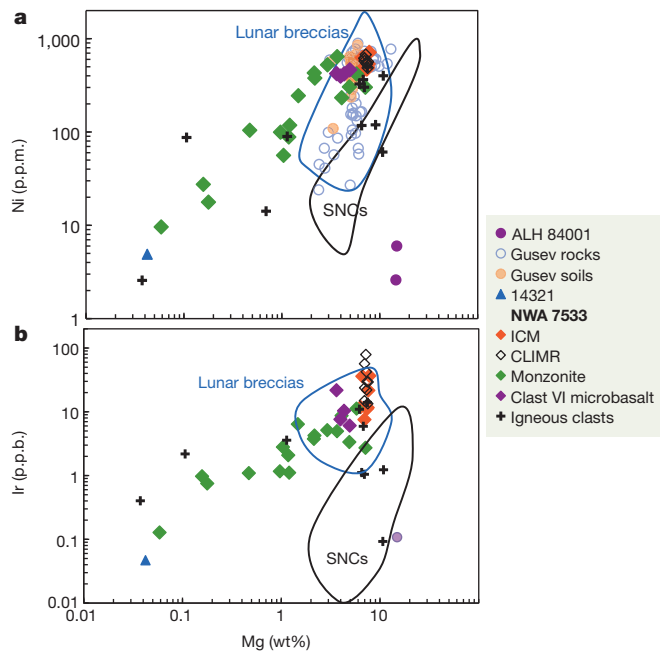
Figure 2 | Siderophile-element abundances in NWA 7533. a, Ni versus Mg, comparing abundances in NWA 7533 components with those in Gusev rocks and soils[12,13], other Martian meteorites (SNCs[23] and ALH 84001[24,25]), Apollo 15–17 breccias[26–28] and lunar meteorites[8,9], and a lunar felsite, 14321, c4 (ref. 29). **b**, Ir versus Mg for the same samples (excluding Gusev rocks and soils, for which Ir data are not available). For literature sources, see above. Some of the *in situ* analyses from NWA 7533 are higher in Ir than any of the lunar breccias, owing to the influence of Ir-rich nuggets.

Figure 3 | Gusev rock and soil analyses[13] have systematically higher Zn abundances than both Martian meteorites and NWA 7533. Pyroxene-rich nakhlites and ALH 84001 are higher in Zn than are olivine-rich chassignites, but none of the known nakhlites is as Fe-rich as some of the igneous-textured clasts from NWA 7533, which extend beyond the SNC field to higher Fe. Together with S and Cl, Zn abundances are systematically enriched in modern soils relative to NWA 7533, presumably because of the lack of liquid water on modern Mars.

crystallized from impact melts enriched in siderophile elements to concentrations similar to those in the ICM and CLIMR.

The remarkable chemical similarity between NWA 7034[3] and Gusev rock and soil analyses[12,13] is confirmed here for NWA 7533. Abundances of major elements (Si, Al, Fe, Ca and Na) in CLIMR and ICM are identical to those in Gusev soils, except for higher Mg in CLIMR and ICM (Supplementary Fig. 4). Among minor elements, the similarity of CLIMR and ICM to Gusev soils is evident in Ni (Fig. 2), Ti and K, although P is up to twice as high as in Gusev soils owing to the abundant chlorapatite in NWA 7533. The fine-grained textures and uniform chemical composition of CLIMR and ICM, which resembles the ubiquitous soil composition reported by NASA's Viking[14], Pathfinder[15] and Mars Exploration Rover[12,13] missions, indicate that these materials contain important amounts of wind-blown dust. Because Ni is a reliable tracer for soil, Gusev rocks with high Ni contents[12,13] may be lithified sediments or impact breccias and cannot be regarded as basalts[10]. Unlike modern Martian soils[11–15], ICM and CLIMR do not show enrichments of S, Cl and Zn with values similar to SNC meteorites (Fig. 3). These elements are likely to be in water-soluble phases in modern soils and the lack of enrichment observed in NWA 7533 components is probably due to the transportation of these salts into ancient seas or lakes[16] by liquid water present on Mars at the time of formation of ICM and CLIMR.

Rare-earth element (REE) abundances for ICM are identical in pattern to those for CLIMR, indicating that the two fine-grained lithologies in this meteorite are derived from similar precursors (Fig. 4). The REE pattern for ICM and CLIMR in NWA 7533 agrees well with the pattern previously reported for bulk NWA 7034[3], except that our *in situ* analyses are less contaminated with leucocratic clasts that carry a striking negative Eu anomaly (Fig. 4). The absolute enrichment of REE varies from 40 to 46 times the CI chondrite level owing to the ubiquitous presence of 10–100-μm clasts in all the analyses. Some of these clasts contribute small Eu anomalies, in the absence of which the REE

patterns of the CLIMR and ICM from NWA 7533 would be smooth and depleted in heavy REEs.

The chemical composition of Martian wind-blown dust, present as ICM and CLIMR in NWA 7533, should provide clues to the original igneous processes that formed the primary Martian crust. A partial-melting model of a primitive mantle composition for Mars (Supplementary Information) indicates that a ~4% partial melt of a fertile mantle containing <1% garnet provides a fit to the CLIMR and ICM REE patterns (Fig. 4). The exact value of the melt fraction depends on the absolute REE abundances, which are diluted by the presence of clasts. If this melt were extracted from the entire Martian mantle it would form a uniform global layer 50 km thick, which is incidentally identical to the average thickness of the Martian crust inferred from gravity and topography measurements by NASA's Mars Global Surveyor[7]. The absence of a garnet (majorite) signature argues against formation of this enriched material as the last dregs of a magma ocean. Combined ^{142}Nd–^{143}Nd isotope evidence in shergottites implies that the formation of the enriched and depleted reservoirs on Mars occurred within the first
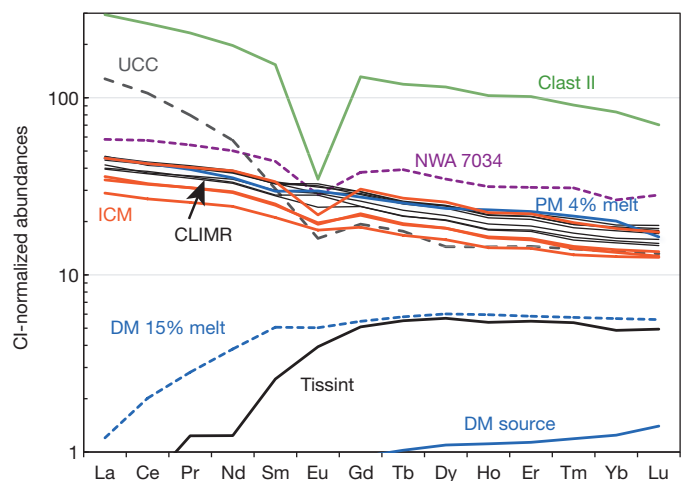
Figure 4 | REE patterns for the representative components of NWA 7533 including the fine-grained ICM and CLIMR. The previously reported bulk REE analysis of NWA 7034[3] (purple) represents a mixture between the ICM or CLIMR and clasts such as monzonite clast II (green). Earth's upper continental crust[30] (UCC) is shown for comparison. The blue curves depict model results: a 4% partial melt of primitive Martian mantle (PM) and the complementary residue termed the depleted Martian source (DM); a higher degree melt (15%) of the DM source; and Tissint[19], a depleted shergottite.

100 Myr of the planet's history[17]. Here we identify the enriched reservoir to be the crust. It is no longer necessary to invoke a magma ocean from Nd isotope evidence for Mars[18], and we take the 100-Myr Nd isotope timescale to imply that the Martian crust formed very early. Removal of this primary melt yields a depleted residue (Fig. 4), which, on subsequent melting ($\sim$15%), yields a composition like that of Tissint meteorite[19], a depleted shergottite. Crustal assimilation by depleted-shergottite magmas then gives rise to intermediate and enriched shergottites[3].

NWA 7533 contains numerous evolved igneous clasts that contain zircons. These evolved lithologies (monzonitic or mugearitic[20] magmas) probably formed by re-melting of the primary Martian crust either at depth in the presence of volatiles[20] or by differentiation of large impact melt sheets. Sensitive high-resolution ion microprobe (SHRIMP) dating of these zircons (Fig. 5) provides a powerful lower limit on the timescale of crustal differentiation. The zircon grains were from 6–70 μm and the spot size for the SHRIMP was $\sim$7 μm in diameter; as a result, some of the analysed spots overlapped the matrix (Supplementary Information and Supplementary Fig. 5). The analyses of overlapping spots were excluded. The analyses for five of ten investigated grains that were entirely within zircon fall on a single discordia line with an upper intercept of 4,428 $\pm$ 25 Myr (1σ) and a lower intercept of 1,712 $\pm$ 85 Myr (1σ) (Fig. 5). The mean squared weighted deviation, of 2.4, most probably results from the analyses being performed on a polished section with some variability in relief, yielding excess scatter of calculated U/Pb in the sample compared with the standard. All analyses, with the exception of two, show $^{206}Pb/^{204}Pb > 400$, with a maximum value of $\sim$1,600 (Supplementary Information). Although these ratios are lower than those usually observed in terrestrial zircons of the same age[6], the common-Pb correction was insensitive to the choice of common-Pb composition. A single, nearly concordant zircon (Z11) with a $^{207}Pb/^{206}Pb$ age of 4114 $\pm$ 30 Myr (1σ) may represent a different age population of zircons in the sample.

These ancient ages for Martian zircons are strikingly similar to the ages of the earliest terrestrial[6] and lunar zircons[5], implying coeval crust formation on the Earth, Moon and Mars. Because the leucocratic clasts formed either by impact or by internal melting of the crust, the events dated by the zircons post-date the emplacement of the Martian crust (4.47 Gyr (ref. 17)) by only $\sim$40 Myr. The cause of the younger age intercept at 1.7 Gyr is not known, but it is close to the Rb–Sr age of 2.1 Gyr for NWA 7034[3], indicating major disturbance of both U–Pb and Rb–Sr ages for the leucocratic clasts at this time.

The combination of compositional and chronological evidence presented here for NWA 7533 implies that it originated from the earliest Martian crust brecciated by impacts. The alkali basalt composition of this crust is now ubiquitously distributed by impacts and wind-blown dust in all major Martian soils sampled by spacecraft landers[11–15]. The observation that Ni remains as high in modern soils as in CLIMR implies minimal subsequent crustal resurfacing on Mars. Evidence for early differentiation ($>$4,400 Myr ago) within the crust to form leucocratic rocks, and the redistribution of these clasts into highland breccias, forms a potent means by which large areas of Martian crust can retain K/Th signatures distinct from that of the uniform wind-blown dust[21]. The early magmatic build-up of the Martian highland crust requires an equally rapid release of volatiles from the Martian interior, forming the early atmosphere and hydrosphere of Mars[1,4], with implications for early Martian climate and biological potential[22]. Further studies of this meteorite will shed light on plutonic rock compositions of the Martian highlands, Martian zirconology and the earliest sedimentary compositions on Mars.

METHODS SUMMARY

Samples of NWA 7533 were analysed using a Tescan VEGA II LSU scanning electron microscope and a Zeiss SIGMA scanning electron microscope at MNHN Paris and ENS Paris, and a CAMECA SX5 electron microprobe at the Université Paris VI. An uncoated section, NWA 7533 section 3, was analysed by laser ablation ICP–MS using an ElectroScientific Instruments New Wave UP193FX ArF excimer (193 nm) laser ablation system coupled to a Thermo Electron Element XR ICP–MS at Florida State University. Altogether, 76 peaks for major and trace elements and their interferences were monitored. Spot sizes of 50–150 μm were used, and the laser repetition rate was 50 Hz, with a fluence of $>$2 GW cm^{-2}. Raster rates were 10 μm s^{-1}. Laser dwell times on a spot were 20 s, resulting in a pit depth of $\sim$100 μm. Relative sensitivity factors obtained from separate standards for many well-characterized lithophile elements agreed to 2–5%, but the accuracy was worse for elements for which only one standard was available, for example NIST SRM 610 ($\sim$10–20%). Before U–Pb analysis, zircons were imaged by cathodoluminescence using a variable-pressure Zeiss EVO scanning electron microscope at Curtin University configured to collect a cathodoluminescence signal, with an acceleration voltage of 10 kV. The working distance was 8.5 mm. Uranium–lead isotope analyses on Au-coated NWA 7533 section 4 were performed on a SHRIMP II at Curtin University under analytical conditions described previously[5]. The beam spot was reduced to 7 μm to effectively analyse the small zircons observed with a primary O^{2-} beam current of 0.5 nA (Methods).

Full Methods and any associated references are available in the online version of the paper.

1. Solomon, S. C. et al. New perspectives on ancient Mars. Science **307**, 1214–1220 (2005).
2. Nimmo, F. & Tanaka, K. Early crustal evolution of Mars. Annu. Rev. Earth Planet. Sci. **33**, 133–161 (2005).
3. Agee, C. B. et al. Unique meteorite from early Amazonian Mars: water-rich basaltic breccia Northwest Africa 7034. Science **339**, 780–785 (2013).
4. Marty, B. & Marti, K. Signatures of early differentiation of Mars. Earth Planet. Sci. Lett. **196**, 251–263 (2002).
5. Nemchin, A. et al. Timing of crystallization of the lunar magma ocean constrained by the oldest zircon. Nature Geosci. **2**, 133–136 (2009).
6. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. Nature **409**, 175–178 (2001).
7. Zuber, M. T. et al. Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. Science **287**, 1788–1793 (2000).
8. Korotev, R. L., Ziegler, R. A., Jolliff, B. L., Irving, A. J. & Bunch, T. E. Compositional and lithological diversity among brecciated lunar meteorites of intermediate iron concentrations. Meteorit. Planet. Sci. **44**, 1287–1322 (2009).
9. Warren, P. H., Ulff-Møller, F. & Kallemeyn, G. W. "New" lunar meteorites: impact melt and regolith breccias and large-scale heterogeneities of the upper lunar crust. Meteorit. Planet. Sci. **40**, 989–1014 (2005).
10. Tuff, J., Wade, J. & Wood, B. J. Volcanism on Mars controlled by early oxidation of the upper mantle. Nature **498**, 342–345 (2013).
11. Yen, A. S. et al. An integrated view of the chemistry and mineralogy of Martian soils. Nature **436**, 49–54 (2005).
12. McSween, H. Y., Taylor, G. J. & Wyatt, M. B. Elemental composition of the Martian crust. Science **324**, 736–739 (2009).
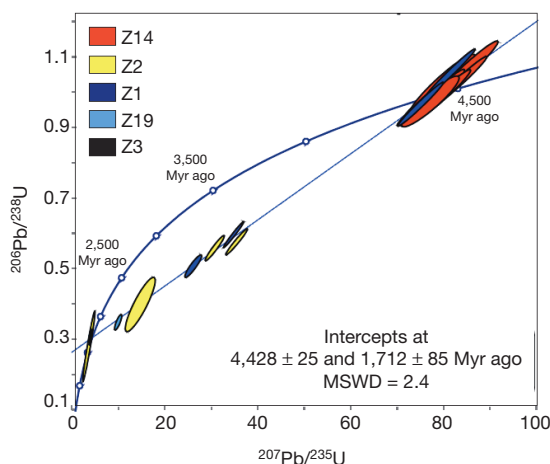
Figure 5 | Concordia plot for SHRIMP analysis of five zircon grains from NWA 7533 section 4 defines a discordia line. Data error ellipses are 2σ. Analyses from three zircons plot on the upper intercept (Z1, Z14, Z15), and the analysis from one grain plots on the lower intercept (Z3). MSWD, mean squared weighted deviation.

13. Gellert, R. *et al.* Alpha Particle X-Ray Spectrometer (APXS): results from Gusev crater and calibration report. *J. Geophys. Res.* **111,** E02S05 (2006).
14. Clark, B. C. *et al.* Chemical composition of Martian fines. *J. Geophys. Res.* **87,** 10,059–10,067 (1982).
15. Foley, C. N., Economou, T. & Clayton, R. N. Final chemical results from the Mars pathfinder alpha proton X-ray spectrometer. *J. Geophys. Res.* **108,** 8096 (2003).
16. Di Achille, G. & Hynek, B. A. Ancient ocean on Mars supported by global distribution of deltas and valleys. *Nature Geosci.* **3,** 459–463 (2010).
17. Debaille, V., Brandon, A. D., Yin, Q.-Z. & Jacobsen, B. Coupled ^{142}Nd-^{143}Nd evidence for a protracted magma ocean in Mars. *Nature* **450,** 525–528 (2007).
18. Elkins-Tanton, L. T. Magma oceans in the inner solar system. *Annu. Rev. Earth Planet. Sci.* **40,** 113–139 (2012).
19. Aoudjehane, H. C. *et al.* Tissint Martian meteorite: a fresh look at the interior, surface, and atmosphere of Mars. *Science* **338,** 785–788 (2012).
20. Stolper, E. M. *et al.* The petrochemistry of Jake_M: a Martian mugearite. *Science* **341,** 1239463 (2013).
21. Taylor, G. J. *et al.* Variations in K/Th on Mars. *J. Geophys. Res.* **111,** E03S06 (2006).
22. Jakosky, B. M. & Phillips, R. J. Mars' volatile and climate history. *Nature* **412,** 237–244 (2001).
23. Brandon, A. D. *et al.* Evolution of the Martian mantle inferred from the ^{187}Re-^{187}Os isotope and highly siderophile element abundance systematics of shergottite meteorites. *Geochim. Cosmochim. Acta* **76,** 206–235 (2012).
24. Warren, P. H., Kallemeyn, G. W. & Kyte, F. T. Origin of planetary cores: Evidence from highly siderophile elements in martian meteorites. *Geochim. Cosmochim. Acta* **63,** 2105–2122 (1999).
25. Kong, P., Ebihara, M. & Palme, H. Siderophile elements in Martian meteorites and implications for core formation in Mars. *Geochim. Cosmochim. Acta* **63,** 1865–1875 (1999).
26. Norman, M. D., Bennett, V. C. & Ryder, G. Targeting the impactors: siderophile element signatures of lunar impact melts from Serenitatis. *Earth Planet. Sci. Lett.* **202,** 217–228 (2002).
27. Palme, H. *et al.* New data on lunar samples and achondrites and a comparison of the least fractionated samples from the Earth, the Moon and the Eucrite parent body. *Proc. Lunar Sci. Conf.* **9,** 25–57 (1978).
28. Wänke, H. *et al.* On the chemistry of lunar samples and achondrites: Primary matter in the lunar highlands. A re-evaluation. *Proc. Lunar Sci. Conf.* **8,** 2191–2213 (1977).
29. Warren, P. H., Taylor, G. J., Keil, K., Shirley, D. N. & Wasson, J. T. Petrology and chemistry of two ''large'' granite clasts from the Moon. *Earth Planet. Sci. Lett.* **64,** 175–185 (1983).
30. Rudnick, R. L. & Gao, S. in *Treatise on Geochemistry* Vol. 3, *The Continental Crust* (ed. Rudnick, R. L.) 1–64 (Elsevier-Pergamon, 2003).

Author Contributions M.H., A.N., B.Z. and R.H.H. had the idea behind and directed the research, and wrote the manuscript. M.H. and B.Z. performed laser ablation ICP–MS analyses at Florida State University; A.N., M.G. and A.K. performed the SHRIMP ion probe U–Pb analyses at Curtin University and interpreted the chronology; B.Z. and C.F. prepared polished samples; R.H.H. and B.Z. performed petrological studies; J.-P.L. and S.P. investigated the mineralogy of the sulphide phases and searched for the carriers of platinum-group elements; C.G. provided separated CLIMR clasts; S.P., D.D., J.-P.L. and B.Z. located and imaged zircon and baddeleyite by scanning electron microscopy.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H. (humayun@magnet.fsu.edu).

METHODS

Laser ablation ICP–MS measurements of NWA 7533. An uncoated section, NWA 7533 section 3, was analysed by laser ablation ICP–MS using an Electro-Scientific Instruments New Wave UP193FX ArF excimer (193 nm) laser ablation system coupled to a Thermo Electron Element XR ICP–MS at Florida State University, as described elsewhere[31–33]. Spot sizes of 50–150 μm were used, the laser repetition rate was 50 Hz and the fluence was >2 GW cm^{-2}. Raster rates were 10 μm s^{-1}. Laser dwell times on a spot were 20 s, resulting in a pit depth of ~100 μm. Altogether, 76 peaks for major and trace elements and their interferences were monitored, and the intensities converted to concentrations using a combination of silicate, sulphide and metal standards, including NIST SRM 610 glass[34]; USGS glasses BHVO-2g, BCR-2g and BIR-1g; NIST SRM 1263a steel[35]; Hoba[36] (IVB); North Chile (Filomena, IIA); and a pyrite crystal. The MPI-DING glasses were measured as independent controls. Major elements were determined using published methods[37]. Relative sensitivity factors obtained from separate standards for many well-characterized lithophile elements agreed to 2–5%, but the accuracy is worse for elements for which only one standard was available, for example NIST SRM 610 (~10–20%). Interference corrections for doubly charged Ba, Nd and Sm ions on Zn, Ga, Ge, As and Se were performed by monitoring ^{137}Ba^{2+}, ^{145}Nd^{2+} and ^{149}Sm^{2+}. Owing to interference from ZrO$^+$ and MoO$^+$, no data are reported for Pd, Ag and Cd here. The absence of suitable standards prevented data from being obtained for Br, I and Hg. Representative chemical compositions for selected samples discussed in the text, peaks monitored and detection limits determined on MPI-DING glasses are provided in Supplementary Table 1, together with the bulk composition of NWA 7034[3]. Section 3 was then carbon-coated and examined by EMP in Paris. Examples of post-ablation images are provided in Supplementary Fig. 3.

SHRIMP U–Pb analyses of zircon and baddeleyite. Before U–Pb analysis, zircons were imaged by cathodoluminescence using a variable-pressure Zeiss EVO scanning electron microscope at Curtin University configured to collect a cathodoluminescence signal, with an acceleration voltage of 10 kV. The working distance was 8.5 mm. Uranium–lead isotope analyses on Au-coated NWA 7533 section 4 were performed on a SHRIMP II at Curtin University under analytical conditions described previously[38–40]. The beam spot was reduced to 7 μm using a 30-μm Kohler aperture to effectively analyse the small zircons observed with a primary O^{2-} beam current of 0.5 nA. Secondary ions were passed to the mass spectrometer operating at a mass resolution ($M/\Delta M$ at 1%) of ~5,000. Each analysis was preceded by a 2-min raster to remove the Au coating and surface contamination. The peak-hopping U–Pb data collection routine consisted of seven scans through the mass stations, with signals measured using an ion-counting electron multiplier. Compared with a typical zircon ion probe analysis, counting times were increased for ^{204}Pb (to 20 s), ^{206}Pb (to 20 s) and ^{207}Pb (to 50 s) to increase the precision of ^{207}Pb/^{206}Pb for individual spot analyses. The sensitivity of the instrument during the session was determined to be 20 c.p.s. p.p.m.$^{-1}$ nA^{-1} using Pb isotopes. Measured Pb/U and Pb/Th ratios in zircon grains were corrected using a 562-Myr-old CZ3 zircon standard[41]. Twenty seven analyses of this standard made during the session resulted in an external error of 2.4% (1σ) in ^{206}Pb/^{238}U, which was added to the errors in ^{206}Pb/^{238}U obtained for each Martian zircon.

Considering that SHRIMP analyses of U/Pb in baddeleyite suffer from strong orientation effects, preventing reliable estimates of U/Pb (ref. 42), only ^{207}Pb/^{206}Pb ages have been calculated for three baddeleyite grains identified in the section (Supplementary Fig. 6). Common Pb in both zircon and baddeleyite was corrected using present-day terrestrial ratios[43], following the observation that much of the common Pb in sections of extraterrestrial materials comes from contamination of the samples during their preparation[44]. However, correcting all analyses using more primitive Pb isotope compositions does not result in any meaningful change in the calculated ages. Raw data have been reduced using SQUID2[45]. Concordia diagrams and intercept calculations were made using Excel add-in ISOPLOT3.75[46]. The calculated data are presented in the Supplementary Table 2 with errors reported at the 1σ level. Ellipses and error bars in all diagrams are shown at the 2σ level and intercept ages are calculated at the 95% confidence level.

31. Humayun, M., Simon, S. B. & Grossman, L. Tungsten and hafnium distribution in calcium-aluminum inclusions (CAIs) from Allende and Efremovka. *Geochim. Cosmochim. Acta* **71,** 4609–4627 (2007).
32. Gaboardi, M. & Humayun, M. Elemental fractionation during LA-ICP-MS analysis of silicate glasses: implications for matrix-independent standardization. *J. Anal. Atomic Spectrom.* **24,** 1188–1197 (2009).
33. Humayun, M. Chondrule cooling rates inferred from diffusive profiles in metal lumps from the Acfer 097 CR2 chondrite. *Meteor. Planet. Sci.* **47,** 1191–1208 (2012).
34. Jochum, K. P. *et al.* Determination of reference values for NIST SRM 610-617 glasses following ISO guidelines. *Geostand. Geoanalyt. Res.* **35,** 397–429 (2011).
35. Campbell, A. J., Humayun, M. & Weisberg, M. K. Siderophile element constraints on the formation of metal in the metal-rich chondrites Bencubbin, Weatherford, and Gujba. *Geochim. Cosmochim. Acta* **66,** 647–660 (2002).
36. Walker, R. J. *et al.* Modeling fractional crystallization of group IVB iron meteorites. *Geochim. Cosmochim. Acta* **72,** 2198–2216 (2008).
37. Humayun, M., Davis, F. A. & Hirschmann, M. M. Major element analysis of natural silicates by laser ablation ICP-MS. *J. Anal. Atomic Spectrom.* **25,** 998–1005 (2010).
38. Compston, W., Williams, I. S. & Meyer, C. U–Pb geochronology of zircons from Lunar Breccia 73217 using a sensitive high mass-resolution ion microprobe. *J. Geophys. Res.* **89,** 525–534 (1984).
39. Nelson, D. R. Compilation of SHRIMP U-Pb geochronology data, 1996. *Geol. Surv. West. Aust. Rec.* **1997/ 2,** 1–11 (1997).
40. Williams, I. S. in *Applications of Microanalytical Techniques to Understanding Mineralising Processes* (eds McKibben, M. A., Shanks, W. C. & Riley, W. I.) 1–35 (Rev. Econ. Geol. 7, Society of Economic Geologists, 1998).
41. Pidgeon, R. T., Furfaro, D., Kennedy, A. K., Nemchin, A. A. & van Bronswjk, W. Calibration of zircon standards for the Curtin SHRIMP. *US Geol. Surv. Circ.* **1107,** 251 (1994).
42. Wingate, M. T. D. & Compston, W. Crystal orientation effects during ion microprobe U-Pb analysis of baddeleyite. *Chem. Geol.* **168,** 75–97 (2000).
43. Stacey, J. S. & Kramers, J. D. Approximation of terrestrial lead isotope evolution by a two-stage model. *Earth Planet. Sci. Lett.* **26,** 207–221 (1975).
44. Nemchin, A. A., Pidgeon, R. T., Whitehouse, M. J., Vaughan, J. P. & Meyer, C. SIMS U-Pb study of zircon from Apollo 14 and 17 breccias: implications for the evolution of lunar KREEP. *Geochim. Cosmochim. Acta* **72,** 668–689 (2008).
45. Ludwig, K. R. *User's Manual for Isoplot 3.60: A Geochronological Toolkit for Microsoft Excel.* Spec. Publ. 4 (Berkeley Geochronological Center, 2008).
46. Ludwig, K. R. *Squid 2 – A User's Manual (rev 2.50).* Spec. Publ. 4 (Berkeley Geochronology Center, 2009).

LETTER

Self-reinforcing impacts of plant invasions change over time

Stephanie G. Yelenik[1]† & Carla M. D'Antonio[1]

Returning native species to habitats degraded by biological invasions is a critical conservation goal[1]. A leading hypothesis poses that exotic plant dominance is self-reinforced by impacts on ecosystem processes, leading to persistent stable states[2–6]. Invaders have been documented to modify fire regimes, alter soil nutrients or shift microbial communities in ways that feed back to benefit themselves over competitors[2,5–7]. However, few studies have followed invasions through time to ask whether ecosystem impacts and feedbacks persist[8,9]. Here we return to woodland sites in Hawai'i Volcanoes National Park that were invaded by exotic C_4 grasses in the 1960s, the ecosystem impacts of which were studied intensively in the 1990s[10–12]. We show that positive feedbacks between exotic grasses and soil nitrogen cycling have broken down, but rather than facilitating native vegetation, the weakening feedbacks facilitate new exotic species. Data from the 1990s showed that exotic grasses increased nitrogen-mineralization rates by two- to fourfold, but were nitrogen-limited[10,12,13]. Thus, the impacts of the invader created a positive feedback early in the invasion. We now show that annual net soil nitrogen mineralization has since dropped to pre-invasion levels. In addition, a seedling outplanting experiment that varied soil nitrogen and grass competition demonstrates that the changing impacts of grasses do not favour native species re-establishment. Instead, decreased nitrogen availability most benefits another aggressive invader, the nitrogen-fixing tree *Morella faya*. Long-term studies of invasions may reveal that ecosystem impacts and feedbacks shift over time, but that this may not benefit native species recovery.

Invasive species have come to the forefront of the conservation movement because of the considerable impact they have on ecosystem composition and functioning, including their impact on threatened and endangered species[14]. In addition to direct competitive effects, invasive species alter disturbance regimes, hydrologic cycles, soil erosion, productivity and nutrient dynamics[2,15–18].

Species that alter ecosystem processes are of special concern for conservation because they alter the rules of the game for resident species[2]. Such changes in ecosystem function are often proposed to feedback positively on the initial invaders by establishing conditions that promote or maintain dominance[2,4–7]. For example, exotic plants that increase soil nitrogen (N) by producing large quantities of nutrient-rich litter may achieve higher growth rates in fertile soils[7,17]. If high soil N is more favourable to exotic than to native species, exotic plants are reinforcing themselves via their ecosystem effects[7,19]. Positive feedbacks can lead to alternative stable states, in which exotic-dominated species assemblages are persistent owing to internal reinforcement[20–22]. Because degraded ecosystems are difficult to restore, the existence of internal feedbacks has become a widely accepted explanation for seeming stability, despite a paucity of evidence.

Although ecosystem impacts of exotic species are commonly studied, there is surprisingly little long-term work in field settings[8,20]. This is an important missing link in our understanding of invader impacts: if ecosystem impacts change over the course of invasion, this may lead to incorrectly predicting invasion outcomes such as alternative states, and result in misdirected management strategies[20]. Ecosystems in alternative stable states typically need large shifts in community composition or environmental conditions to be restored[20,22], yet some invader effects might not be expected to be stable in the long term. For example, as resources such as soil N increase, it is probable that other resources will become limiting (for example, light or phosphorus)[23], changing the relative benefit derived from the invader's enhancement of N cycling. Indeed, some observational studies have shown that seemingly stable invasions have become less robust over time, giving way to successional progression[24,25]. Understanding when and why invasions are self-reinforcing will guide management and inform ecological theory.

In Hawai'i, invasions by exotic C_4 perennial grasses have considerably altered plant community composition and ecosystem processes in seasonally dry woodlands dominated by the native tree *Metrosideros polymorpha*. In the 50 years since widespread invasion, exotic grass species have increased both fire frequency and size in Hawai'i Volcanoes National Park, leading to local declines in native species and loss of net primary productivity[16,26]. Dominance of the exotic grass *Melinis minutiflora* in the initial decade after fire (1988–1998) was associated with increased (2–4-fold) annual N-mineralization rates compared to unburned native woodland[16], whereas N-fertilization studies showed strong N responsiveness of the grasses[13]. Accelerated N cycling, in combination with N limitation of grasses, appeared to contribute to a positive feedback facilitating exotic grass dominance.

Here, we return to Hawai'i Volcanoes National Park to repeat measures of nutrient cycling and plant community change. Our data show that, in the last 17 years, N-mineralization rates in sites dominated by the exotic grass *Melinis minutiflora* have declined by half, thereby returning to pre-invasion levels, while rates in native woodland sites have remained constant (Fig. 1a). This reversal of invader impacts is possibly due to a previous mismatch between N availability and N uptake, leading to high potential for soil N loss[16]. Biomass and net primary productivity in grassland were greatly reduced after invasion and fire, such that plants in invaded sites did not take up the quantity of N being mineralized early in *Melinis* invasion. By contrast, native woodland showed similar annual rates of plant N uptake versus net N mineralization[16]. There was also a mismatch between the annual timing of N mineralization and phenology of *Melinis*. A high amount of N mineralization was found to take place in winter when exotic grasses are less active but rainfall is relatively high, potentially leading to N loss through leaching or denitrification[10,16].

It is possible that differences in N mineralization in years 1994–95 versus years 2011–12 were due to differences in rainfall. However, rainfall during the two sampling periods was similar (Extended Data Fig. 1) and rainfall did not correlate with differences between grassland and woodland sites (Extended Data Fig. 2). Finally, recent laboratory N-mineralization incubations with constant soil moisture showed similar patterns to the 2011–12 field data (Extended Data Fig. 3). Taken together, this suggests that changes in the relative difference in N mineralization between *Melinis*-invaded and intact woodland were not simply due to differences in rainfall.

[1]Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106, USA. †Present address: US Geological Survey, Pacific Island Ecosystems Research Center, Hawai'i Volcanoes National Park, Hawai'i 96718, USA.
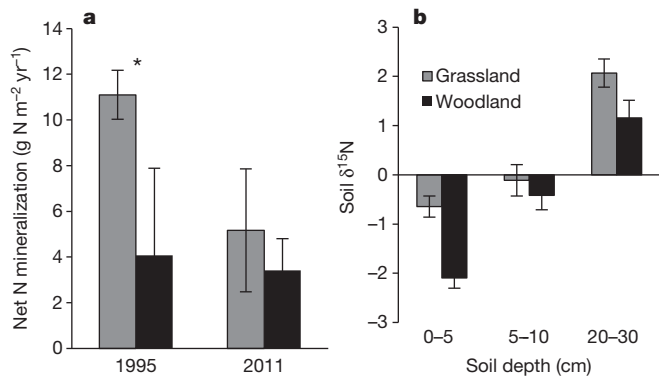
Figure 1 | Ecosystem impacts of *Melinis* invasion over time, and through the soil profile. a, Net N-mineralization rates in 1995 and 2011 in exotic grassland versus native *Metrosideros* woodland sites. 1995 data showed differences[10] ($P < 0.05$), whereas data from present conditions do not (one-way analysis of variance (ANOVA), habitat as fixed effect and site as random effect, $n = 15$: $P = 0.65$). **b**, Soil $\delta^{15}N$ from grassland and woodland sites at varying depths taken during 2011. Grassland soils are consistently less negative, suggesting greater N losses (two-way ANOVA, habitat and soil depth as fixed effects and site as random effect, $n = 5$: habitat, $P < 0.01$; soil depth, $P < 0.01$). Bars represent means ± 1 s.e.

Consistent with the hypothesis that N-cycling rates have declined in the invaded ecosystem, *Melinis* foliar N has decreased by 30%. In 1995, foliar %N was 0.43 ± 0.02, whereas present values average 0.32 ± 0.01. In addition, *Melinis* soil $\delta^{15}N$ values currently have a greater proportion of the heavier isotope than woodland soils (Fig. 1b). This suggests that exotic systems experience greater N losses than woodlands because N lost to pathways such as denitrification is depleted in the heavier isotope[27]. That invasion has lowered ecosystem N supply rates suggests that N limitation of *Melinis* may be exacerbated[13], causing initial positive feedbacks to weaken. Indeed, recent vegetation data show that, concurrent with shifting N mineralization, *Melinis* live cover has decreased[26], and live biomass has decreased from values of 600 g m^{-2} (refs 13, 16) to an average of 413 g m^{-2} during 2011 (Fig. 2). Although positive feedbacks are an important contributor to alterative stable states, many other life history (for example, seed set, growth rate) and stochastic variables have a role in overall community dynamics[9,21,22]. However, this is, to our knowledge, the first study in which a positive feedback that coincides with dominance has been shown to shift to a negative feedback, and decreasing dominance, over time.
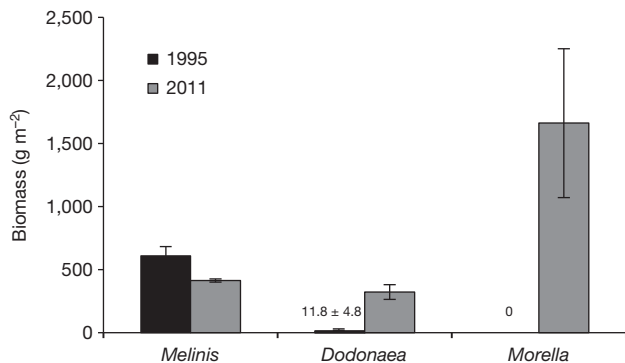
Figure 2 | Changes in dominant species biomass in exotic grassland habitat over time. The exotic grass *Melinis* has decreased in biomass over time, whereas *Dodonaea* and *Morella* have increased since the 1990s. Note that *Morella* biomass has increased more dramatically than *Dodonaea*, potentially owing to its greater response to the changing ecosystem impacts of *Melinis*. Bars are means ± 1 s.e.; all species 2011 biomass, $n = 11$; 1995 *Melinis* biomass[11,13], $n = 5$; 1995 *Dodonaea* biomass[30], $n = 3$. *Morella* was not found on grassland transects in 1995.

If changes in self-reinforcing invader impacts are leading to a loss of *Melinis* dominance, we wanted to know which species would benefit from gaps or declines in *Melinis* cover. We used a large outplanting experiment to test how a suite of regionally available native and exotic woody species responded to shifting ecosystem impacts of *Melinis*. Specifically, we added N (simulating higher soil N in early *Melinis* invasion) and/or clipped aboveground *Melinis* (simulating gaps in cover later in invasion) in a factorial field experiment that included seven species, replicated over seven sites in invaded grasslands. We refer to clipped *Melinis* as 'reduced competition' because intact root systems continued to produce shoots and take up resources. Those outplanted species that receive a greater relative benefit from reduced *Melinis* competition than from added N are favoured later in *Melinis* invasion, and therefore are predicted to be more likely to fill open space in degraded *Melinis* grasslands.

Five out of the seven outplanted species responded similarly to treatments, with growth rates and survivorship increasing from both reduced *Melinis* competition and N additions (Fig. 3). Because there were no differences between treatment effects for these native and exotic species we suggest that the changing ecosystem impacts of *Melinis* have not altered their ability to colonize *Melinis* grasslands. Exceptions to this pattern included the N-fixing trees: native *Acacia koa* and exotic *Morella faya*. These species benefited more from reduced *Melinis* competition than from N addition, suggesting that a change in ecosystem impact would release them from *Melinis* competition to the greatest degree out of the species tested (Fig. 3). It should be noted that other species
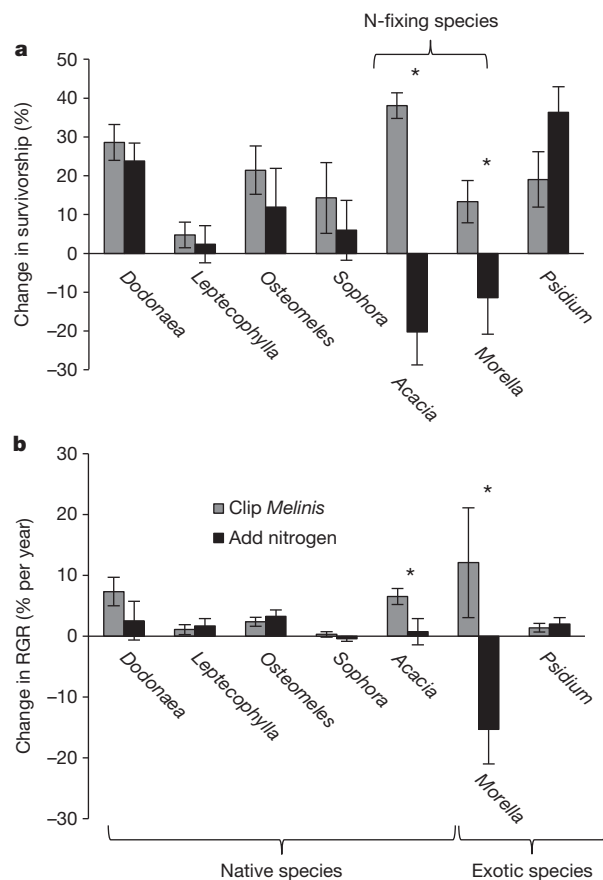
Figure 3 | Assessing the changing impacts of *Melinis* invasion on native and exotic seedlings. a, b, We compared the effects of adding N fertilizer (similar to gaining benefit from increased soil N in early *Melinis* invasion) to clipping *Melinis* (similar to gaining benefit from reduced competition later in *Melinis* invasion) on seedling survival rate (**a**) and RGR in outplantings (**b**). Asterisks show significant differences between clipping and fertilization effects at the $P < 0.05$ level ($n = 7$), and suggest that changing ecosystem impacts of *Melinis* over time will alter growth rates and survivorship of these species. Bars represent means ± 1 s.e. See Methods for statistics.

benefitted from reduced *Melinis*, including the native shrub *Dodonaea*, which also has greater biomass on the landscape than in 1995. However, its increase in biomass is much less than *Morella* over time (Fig. 2), which may be because the benefit received from shifting ecosystem impacts of *Melinis* is not as great for *Dodonaea*.

Taken together, our results suggest that changing *Melinis* impacts are leading to negative feedbacks with N-fixing species in the long term. During early invasion (Fig. 4a), *Melinis* increased soil N availability[10], which, given that it was N limited[13], would be a positive feedback. However, during late invasion (Fig. 4b), *Melinis* is associated with ecosystem N depletion (Fig. 1a), which we have shown benefits N fixers (Fig. 3). At the same time, N fixers have large localized positive effects on soil N pools[18], which benefits grasses more than N fixers, as N-fixing trees respond negatively to N additions (Fig. 3). Therefore, each functional group is ultimately altering soils in ways that benefit the other more than itself. Such negative feedbacks are stabilizing and foster co-existence[7,19], which, all else being equal, allow each functional group to colonize areas dominated by the other.

In fact, the N-fixing tree *Morella* is moving rapidly into exotic grass sites. Although there were none established on permanent transects in the 1990s, this species currently makes up >60% of standing biomass in exotic grasslands (Fig. 2). By contrast, no *Acacia* trees have recruited into study sites, a difference potentially due to dispersal limitation. Whereas *Morella* is bird dispersed and invades widely across Hawai'i, *Acacia* is a heavy seeded, slow disperser[28] that is locally of limited distribution. A trait-based management option would be to facilitate species with the same trait—N fixation—that is allowing *Morella* to increase in abundance in this later *Melinis* invasion stage[29]. Aggressively outplanting *Acacia* would overcome dispersal limitation, allowing it to pre-empt resources before *Morella* arrival, although whether or not this novel ecosystem is desirable should be explored with managers. Although *Sophora* is also a native N-fixing tree, its low relative growth rates (RGRs; Extended Data Fig. 4) makes it an imperfect restoration species, especially if a goal is to outcompete the non-native *Morella*.

Understanding how plant invasions alter ecosystems in the long term, and what this means for community trajectories, is critical for informing restoration practices[3,8,15]. We offer what is to our knowledge the first long-term study of invader feedbacks, and show that they weaken over time. Although this facilitates community succession away from the initial invader, our data suggest that without further management, native species may not gain the advantage of altered invader impacts. Taking a mechanistic perspective to feedbacks and stable states in more case studies will help ecologists to gain a general understanding of when feedbacks can be predicted to be persistent in the long term. For example,

feedbacks that are created by the effects of a single species on soil nutrient cycling might break down as other nutrients become more limiting. Conversely, feedbacks with fire may be more persistent, as they are consistently reset, and thus in a state of constant disequilibrium. Long-term studies of invader feedbacks are needed to test these important ecological ideas.

METHODS SUMMARY

Native unburned *Metrosideros* woodland and burned exotic grassland habitats, corresponding to those used for collecting previous data, were located in Hawai'i Volcanoes National Park[11–14,17,26]. We sampled soil and foliage from three replicate locations within five sites in each habitat type between 2010 and 2012 to compare to 1990s data. We used intact soil cores to measure net N mineralization using protocols from 1995 (ref. 17). These were repeated bimonthly over 1 year in both habitats. *Melinis* foliage for %N was taken from fully expanded green leaf material. Soils were cored to 5, 10 and 30 cm to explore changes in δ^{15}N, informing ecosystem N loss. Woody biomass densities were measured in 11 randomly chosen 10×10 m plots in the exotic grassland habitat. All woody plants were measured for height and basal diameter and allometric equations obtained by harvesting 6–20 measured shrubs per species. Data for 1990s *Dodonaea* biomass were obtained from ref. 30. *Melinis* biomass was estimated in the 1990s by harvesting one 4 m^2 plot of grass for each of the five sites in summer and winter and averaging values across seasons. Biomass for 2010–11 was obtained by harvesting three 0.25 m^2 quadrats for each of the five sites with 100% live *Melinis* cover on seven dates between October 2010 and December 2011. We then multiplied the average live biomass values at 100% cover times the actual live per cent cover of *Melinis* censused in 20 1 m^2 subplots within each 10×10 m woody plant density plot. Finally, we established an outplanting experiment in exotic grasslands exploring the differential effects of *Melinis* invasion over time, using a fully factorial design of added N and reduced *Melinis* competition, in December 2011. We used seven species (native and exotic) found locally, planted in monoculture. Seedlings were measured for relative growth rates and survivorship after 8 months. See Methods for detailed protocols.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Dobson, A. P., Bradshaw, A. D. & Baker, A. J. M. Hopes for the future: restoration ecology and conservation biology. *Science* **277,** 515–522 (1997).
2. D'Antonio, C. M. & Vitousek, P. M. Biological invasions by exotic grasses, the grass fire cycle, and global change. *Annu. Rev. Ecol. Syst.* **23,** 63–87 (1992).
3. Suding, K. N., Gross, K. L. & Houseman, G. R. Alternative states and positive feedbacks in restoration ecology. *Trends Ecol. Evol.* **19,** 46–53 (2004).
4. Callaway, R. M., Thelen, G. C., Rodriguez, A. & Holben, W. E. Soil biota and exotic plant invasion. *Nature* **427,** 731–733 (2004).
5. Kulmatiski, A., Beard, K. H., Stevens, J. R. & Cobbold, S. M. Plant–soil feedbacks: a meta-analytical review. *Ecol. Lett.* **11,** 980–992 (2008).
6. van der Putten, W. H., Klironomos, J. N. & Wardle, D. A. Microbial ecology of biological invasions. *ISME J.* **1,** 28–37 (2007).
7. Yelenik, S. G. & Levine, J. M. The role of plant–soil feedbacks in driving native species recovery. *Ecology* **92,** 66–74 (2011).
8. Strayer, D. L., Eviner, V. T., Jeschke, J. M. & Pace, M. L. Understanding the long-term effects of species invasions. *Trends Ecol. Evol.* **21,** 645–651 (2006).
9. Levine, J. M., Pachepsky, E., Kendall, B. E., Yelenik, S. G. & Lambers, J. H. Plant–soil feedbacks and invasive spread. *Ecol. Lett.* **9,** 1005–1014 (2006).
10. Mack, M. C. & D'Antonio, C. M. Exotic grasses alter controls over soil nitrogen dynamics in a Hawaiian woodland. *Ecol. Appl.* **13,** 154–166 (2003).
11. D'Antonio, C. M., Hughes, R. F. & Vitousek, P. M. Factors influencing dynamics of two invasive C$_4$ grasses in seasonally dry Hawaiian woodlands. *Ecology* **82,** 89–104 (2001).
12. Hughes, F., Vitousek, P. M. & Tunison, T. Alien grass invasion and fire in the seasonal submontane zone of Hawaii. *Ecology* **72,** 743–746 (1991).
13. D'Antonio, C. M. & Mack, M. C. Nutrient limitation in a fire-derived, nitrogen-rich Hawaiian grassland. *Biotropica* **38,** 458–467 (2006).
14. Pimentel, D., Zuniga, R. & Morrison, D. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* **52,** 273–288 (2005).
15. Simberloff, D. Invasional meltdown 6 years later: important phenomenon, unfortunate metaphor, or both? *Ecol. Lett.* **9,** 912–919 (2006).
16. Mack, M. C., D'Antonio, C. M. & Ley, R. E. Alteration of ecosystem nitrogen dynamics by exotic plants: a case study of C$_4$ grasses in Hawaii. *Ecol. Appl.* **11,** 1323–1335 (2001).
17. Allison, S. D. & Vitousek, P. M. Rapid nutrient cycling in leaf litter from invasive plants in Hawai'i. *Oecologia* **141,** 612–619 (2004).

Figure 4 | Feedbacks between *Melinis* and soils change over time, ultimately leading to negative feedbacks with N-fixing species. **a**, Early in invasion, *Melinis* increases soil N cycling[10], increasing its own productivity[13]. **b**, Over time ecosystem N losses deplete soil N (Fig. 1), reducing *Melinis* cover (Fig. 2). Open space is beneficial for N-fixing trees (Fig. 3), which, over time, increase localized soil N pools. In these locations, *Melinis* production is increased[13]. Thus, the effect of each species on localized soil N benefits the other species more than itself, which, all else being equal, promotes co-existence. Figure follows model in ref. 19.

18. Vitousek, P. M. & Walker, L. R. Biological invasion by *Myrica faya* in Hawai'i: plant demography, nitrogen fixation, ecosystem effects. *Ecol. Monogr.* **59,** 247–265 (1989).
19. Bever, J. D., Westover, K. M. & Antonovics, J. Incorporating the soil community into plant population dynamics: the utility of the feedback approach. *J. Ecol.* **85,** 561–573 (1997).
20. Suding, K. N. & Hobbs, R. J. Threshold models in restoration and conservation: a developing framework. *Trends Ecol. Evol.* **24,** 271–279 (2009).
21. Beisner, B. E., Haydon, D. T. & Cuddington, K. Alternative stable states in ecology. *Front. Ecol. Environ.* **1,** 376–382 (2003).
22. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413,** 591–596 (2001).
23. Tilman, D. Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecol. Monogr.* **57,** 189–214 (1987).
24. Sullivan, J. J., Williams, P. A. & Timmins, S. M. Secondary forest succession differs through naturalised gorse and native kanuka near Wellington and Nelson. *N. Z. J. Ecol.* **31,** 22–38 (2007).
25. Beatty, S. W. & Licari, D. L. Invasion of fennel (*Foeniculum vulgare*) into shrub communities on Santa Cruz Island, California. *Madroño* **39,** 54–66 (1992).
26. D'Antonio, C. M., Hughes, R. F. & Tunison, J. T. Long-term impacts of invasive grasses and subsequent fire in seasonally dry Hawaiian woodlands. *Ecol. Appl.* **21,** 1617–1628 (2011).
27. Nadelhoffer, K. J. & Fry, B. in *Stable Isotopes in Ecology and Environmental Science* Vol. 316 (eds Lajtha, K. & Michener, R. M.) 22–44 (Blackwell, 1994).
28. Scowcroft, P. G. Parent tree effects on reestablishment of *Acacia koa* in abandoned pasture and the influence of initial density on stand development. *New Forests* **44,** 409–426 (2013).
29. Funk, J. L., Cleland, E. E., Suding, K. N. & Zavaleta, E. S. Restoration through reassembly: plant traits and invasion resistance. *Trends Ecol. Evol.* **23,** 695–703 (2008).
30. Mack, M. C. *Effects of Ecotic Grass Invasion on Ecosystem Nitrogen Dynamics in a Hawaiian Woodland.* PhD thesis, Univ. California, Berkeley (1998).

Author Contributions S.G.Y. and C.M.D. conceived and designed the study, managed the project, performed laboratory and field work and edited the manuscript. S.G.Y. analysed the data, developed the figures and drafted the initial manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G.Y. (syelenik@usgs.gov).

METHODS

Sites. Native unburned *Metrosideros* woodland and burned exotic grassland habitats, which corresponded to those used for collecting previous data[10–12,16,26], were located in Hawai'i Volcanoes National Park. We took soil and foliar samples from three replicate locations within five sites in each habitat (exotic grassland versus native *Metrosideros* woodland) type. These five sites per habitat are adjacent to the five transect locations studied in all of the previous published work[10–12,16,26].

Soil N cycling. We used intact core soil N-mineralization protocols from 1995 to quantify net N mineralization under field conditions[16], which were replicated seven times over 14 months starting in October 2010. Although intact cores were located at random points within sites, in the grassland habitat we avoided pig-disturbed, bare areas, or entirely dead patches of *Melinis* as well as *M. faya* individuals. Random points were selected along a central transect through the centre of each site. Then at each of the three random points, we flipped a coin as to whether to core on the left or right side of the transect tape and a second random number was generated between 0 and 5 m to place the core location at a random distance from the transect. We also conducted laboratory assays of potential net N mineralization to assess rates without the confounding effects of different microclimates. For these we took two randomly placed (as described previously) 10-cm-deep cores at each site, sieved to 2 mm, and held the soil at constant moisture (70% water-holding capacity) and temperature (23 °C). We extracted the t_0 soils 24 h post wet up and the t_1 soils 30 days later.

Soil for ^{15}N analysis was taken using a 3.8-cm diameter core to 5-, 10- and 30-cm depth in both January and July 2012 at random locations within sites as described above. All soil was air dried and ground with a mortar and pestle. Data were identical between time points, and so only July is presented. *Melinis* foliage for %N was taken from fully expanded green leaf material in November 2010 and July 2011, dried and ground with a ball mill. To be consistent with 1990s timing we present July data here but values were similar. Foliar C:N and ^{15}N samples were analysed at the University of Hawai'i, Hilo Analytical Laboratory using a Costech ECS CHNSO Analyzer (Costech Analytical Technologies), and inorganic N in soil extracts were analysed at UC Santa Barbara with a Lachat flow-injection auto analyser (Lachat Instruments). For all soils data, outliers ($\geq$ mean $\pm$ 2 s.d.) were discarded from the analyses. Data were tested for normality to assure that assumptions of parametric tests were met. See figure legends for statistics.

Plant biomass. Woody biomass densities were measured in 11 randomly chosen locations near the end points of the original transects studied in the 1990s[12,26]. The sampled subplots were 10×10 m. All woody plants were measured for height and basal diameter and allometric equations obtained by harvesting 6–20 measured shrubs per species. Data for 1990s *Dodonaea* biomass were obtained from ref. 30. *Melinis* biomass was estimated in the 1990s by harvesting one 4 m² plot of grass for each of the five sites in summer and winter, separating it into live versus dead plant material, and drying subsamples to correct for field moist weights. Values were

averaged across seasons. Biomass for 2010–11 was obtained by harvesting three 0.25 m² quadrats for each of the five sites with 100% live *Melinis* cover on seven dates between October 2010 and December 2011. Quadrat locations were random except to avoid *M. faya*. Biomass was separated into live versus dead and dried. We then multiplied the average live biomass values at 100% cover times the actual live per cent cover of *Melinis* censused in 20 random 1 m² subplots within each 10×10 m woody plant density plot. This method probably overestimates *Melinis* biomass because of the larger edge to interior ratio for the small harvest plots compared to those harvested in the 1990s, but the values were still lower. Thus, our results of declining *Melinis* biomass and cover over time may be conservative.

Outplanting experiment. We established an outplanting experiment in which we explored the differential effects of *Melinis* invasion over time: higher soil N, representative of early invasion, and lower *Melinis* competition, representative of later *Melinis* invasion. We used seven replicate sites across the invaded grassland habitat to establish the outplant experiment using a fully factorial design with N fertilization (10 g N m^{-2} as urea, half added September 2011, and half added January 2012) and competition removal (clipping *Melinis* at soil surface) treatments. These seven replicate sites were chosen to be at least 100 m apart, away from *M. faya*, with soil at least 30 cm deep, and not a part of the soil N-mineralization core sampling areas. Seedling species, which were the most common in the local species pool (Extended Data Fig. 4), were grown from local seed for 5 months before outplanting in December 2011 at a density of six individuals per 0.25 m² monoculture plot. We planted multiple individuals (six) per species per site to account for mortality, which can be high in this dry ecosystem, and to track survivorship. We separated seedlings into three size classes and included equal numbers from each size class in each replicate to control for initial seedling size effects. Seedlings were small and planted far enough apart, and remained small enough, that we do not feel that they experienced intraspecific competition.

We measured seedlings for initial height and width, re-measured at 8 months, and calculated RGRs by modelling them as inverted cones. We used the average RGR of the initial six seedlings per monoculture plot to compare treatment effects (Extended Data Fig. 4). To calculate change in RGR (Fig. 3), we subtracted the average RGR in control plots (for example, no N fertilizer added in clipped and non-clipped plots) from RGR in treatment plots (for example, N added in clipped and non-clipped plots). We used one-way ANOVAs ($n = 7$) to test for treatment differences (adding N versus clipping *Melinis*). Although all data were normally distributed, we did find unequal variances for some species. However, non-parametric Kruskal–Wallis tests, which do not assume equal variances, showed similar results (that is, the same species showed significant differences between treatments). Survivorship data were percentages based on the number surviving of the initial six seedlings. We therefore used logistic regression to compare treatments for each species ($n = 7$). For outplanting data, outliers (mean $\pm$ 2 s.d.) were discarded from the analyses, which resulted in removing one data point from the growth rate data.
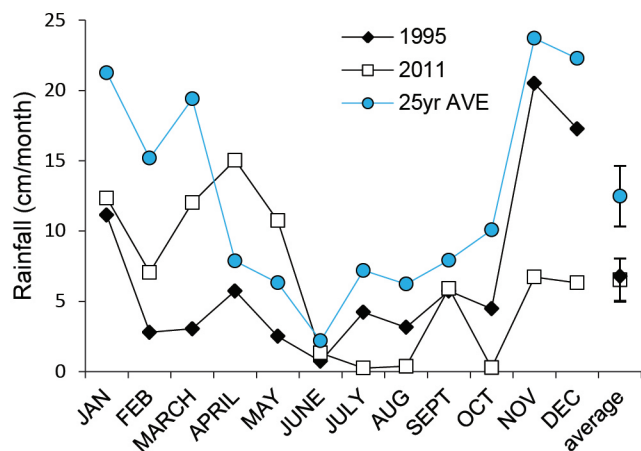
Extended Data Figure 1 | Monthly rainfall over the study periods, and the 25-year average monthly rainfall. Monthly rainfall over the course of 1 year during the 1995 and 2011 year-long sampling periods for net N mineralization (Fig. 1a). The last point in the series shows the average monthly rainfall for that year (points are means ± 1 s.e.). Also shown in blue is the same data for the 25-year rainfall average. Note that 1995 and 2011 have similar rainfall on average over the year, approximately 45% lower than the 25-year rainfall average.
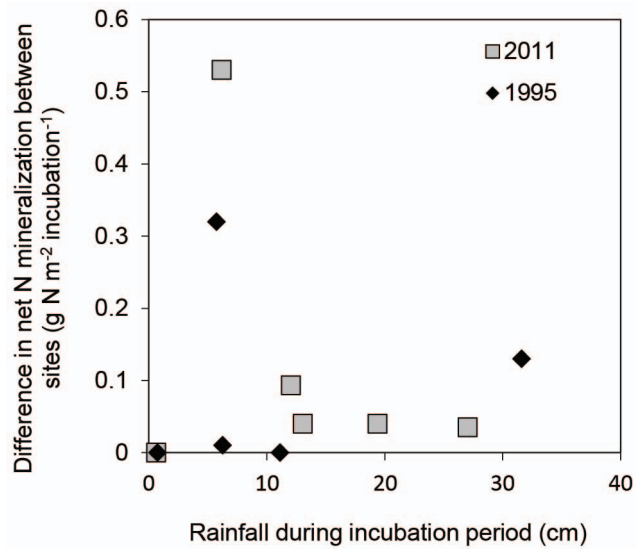
Extended Data Figure 2 | Relationship between net N mineralization and rainfall over the study periods. Differences in net N mineralization between exotic grassland and native *Metrosideros* woodland sites in relation to monthly rainfall for the 1994–95 and 2011–12 study periods. The lack of relationship ($r^2 = 0.01$, $P = 0.74$, $n = 11$) between site differences and monthly rainfall suggests that rainfall did not drive patterns in net N mineralization, or the relationship between invaded and intact woodland sites (Fig. 1a).
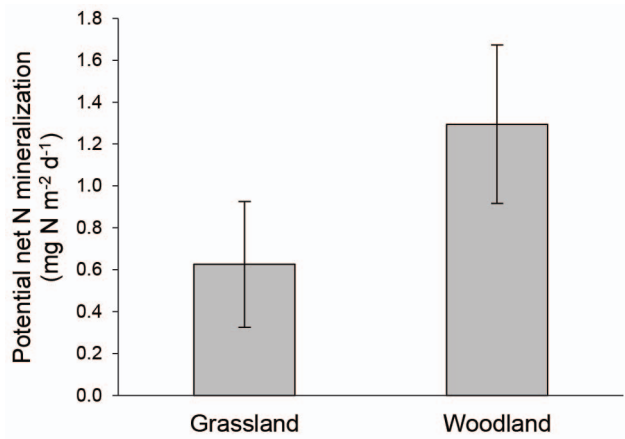
Extended Data Figure 3 | Potential net N mineralization from laboratory assays. Net N-mineralization incubations from the laboratory, where soils were held at 70% water-holding capacity and 23 °C. That there was no difference between exotic grassland and native woodland habitats (one-way ANOVA, habitat as fixed effect: $P = 0.19$, $n = 10$) matches results from intact field cores (Fig. 1a), suggesting that differences in climate between sites, which may have varied in the field, did not alter general results for net N mineralization. Bars represent means $\pm$ 1 s.e. We also ran the analysis with a Kruskal–Wallis test to account for unequal variances, although results were similar ($P = 0.43$, $n = 10$).
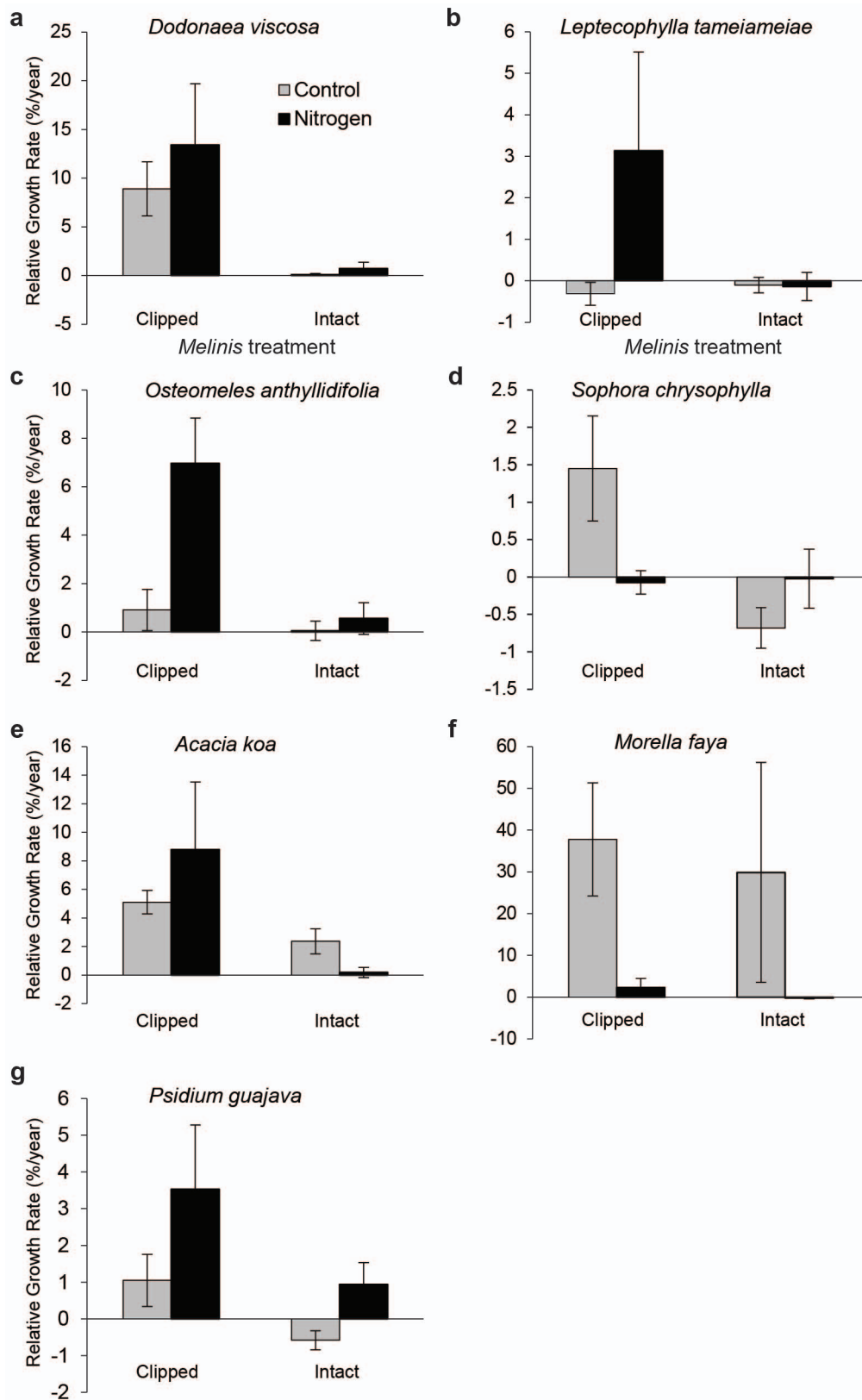
Extended Data Figure 4 | RGRs for seedlings in the outplanting experiment. RGRs were calculated after 8 months for the native seedlings (**a–e**) and the exotic seedlings (**f, g**). **a**, *Dodonaea viscosa* ('ā'ali'i). **b**, *Leptecophylla tameiameiae* (pūkiawe). **c**, *Osteomeles anthyllidifolia* ('ūlei). **d**, *Sophora chrysophylla* (māmane). **e**, *Acacia koa* (koa). **f**, *Morella faya* (faya). **g**, *Psidium guajava* (guava). Bars represent means ± 1 s.e.

Cortical interneurons that specialize in disinhibitory control

Hyun-Jae Pi[1], Balázs Hangya[1,2], Duda Kvitsiani[1], Joshua I. Sanders[1], Z. Josh Huang[1] & Adam Kepecs[1]

In the mammalian cerebral cortex the diversity of interneuronal subtypes underlies a division of labour subserving distinct modes of inhibitory control[1–7]. A unique mode of inhibitory control may be provided by inhibitory neurons that specifically suppress the firing of other inhibitory neurons. Such disinhibition could lead to the selective amplification of local processing and serve the important computational functions of gating and gain modulation[8,9]. Although several interneuron populations are known to target other interneurons to varying degrees[10–15], little is known about interneurons specializing in disinhibition and their *in vivo* function. Here we show that a class of interneurons that express vasoactive intestinal polypeptide (VIP) mediates disinhibitory control in multiple areas of neocortex and is recruited by reinforcement signals. By combining optogenetic activation with single-cell recordings, we examined the functional role of VIP interneurons in awake mice, and investigated the underlying circuit mechanisms *in vitro* in auditory and medial prefrontal cortices. We identified a basic disinhibitory circuit module in which activation of VIP interneurons transiently suppresses primarily somatostatin- and a fraction of parvalbumin-expressing inhibitory interneurons that specialize in the control of the input and output of principal cells, respectively[3,6,16,17]. During the performance of an auditory discrimination task, reinforcement signals (reward and punishment) strongly and uniformly activated VIP neurons in auditory cortex, and in turn VIP recruitment increased the gain of a functional subpopulation of principal neurons. These results reveal a specific cell type and microcircuit underlying disinhibitory control in cortex and demonstrate that it is activated under specific behavioural conditions.

Cortical inhibitory interneurons display great diversity in their physiology, connectivity and synaptic dynamics, but it has long been debated whether and to what extent function of an interneuron type follows from a unique combination of these properties[7]. The possibility that different interneuron cell types perform distinct circuit operations holds great promise for unravelling the logic of cortical microcircuits. Nevertheless, little is known about the functional roles of different interneuron subtypes, especially in awake and behaving animals. Multiple populations of interneurons differentially target distinct subregions of pyramidal cells leading to different modes of inhibitory control. Disinhibition of principal neurons mediated by inhibition targeted onto other inhibitory neurons can provide an additional layer of control, generating a powerful computational mechanism for increasing the gain of principal neurons. Recent work identified a population of layer 1 interneurons that mediate disinhibitory control over cortical processing[13,18] and thereby enable associative learning[18]. Previous studies proposed that VIP-expressing interneurons are a candidate cell type specializing in disinhibition because they seem to mainly target other interneurons[10–12,15]. Indeed, VIP expression demarcates a small population of all interneurons (~15%), distinct from the two major interneuron populations defined by parvalbumin (PV; also called PVALB) and somatostatin (SOM) expression[19,20]. However, whether and how VIP interneurons mediate disinhibition

in vivo and when they are recruited during behaviour has remained elusive.

We examined the function of VIP interneurons in two functionally different cortical regions: auditory cortex (ACx) and medial prefrontal cortex (mPFC). Channelrhodopsin-2 (ChR2)[4,5], a light-activated cation channel, was targeted to VIP neurons using a VIP-IRES-Cre[21] knock-in mouse line by either breeding with Ai32 (ref. 22) (ChR2 reporter line) or using viral delivery (Fig. 1a and Extended Data Fig. 1a–d). To explore the function of VIP interneurons in circuit operations, we acquired extracellular recordings in awake mice using miniature microdrives that house an optical fibre and six tetrodes for simultaneous light stimulation and recording (Fig. 1b and Extended Data Fig. 1e, f).

We first characterized the impact of VIP neurons on the cortical network by synchronously activating them using 1-ms light pulses in ACx and mPFC. Brief light stimulation of this sparse population of VIP-expressing interneurons (1–2% of cortical neurons[19,20]) resulted in a disproportionately broad effect, generating significant firing-rate changes in ~20% of cortical cells (ACx, 130 of 495; mPFC, 26 of 155) for tens to hundreds of milliseconds after the pulse. Examination of the light-triggered activity profiles of neurons revealed three distinct groups (Fig. 1c–h). The first group of neurons was strongly activated at a very low latency (ACx, 2.5 ± 0.22 ms; mPFC, 1.3 ± 0.40 ms; mean $\pm$ s.e.m.), with low jitter (ACx, 1.4 ± 0.21; mPFC, 1.1 ± 0.10 ms; mean $\pm$ s.e.m.) and high reliability across trials (ACx, 0.62 ± 0.10 at 0.5 Hz; mPFC, 0.64 ± 0.17 at 10–20 Hz; mean $\pm$ s.e.m.), indicative of direct light activation (Fig. 1c, g, h and Extended Data Fig. 2a, d). Because ChR2 is expressed under the control of the *Vip* promoter (Fig. 1a and Extended Data Fig. 1), we concluded that the directly activated group comprises VIP interneurons. A second group of neurons was inhibited by the photostimulation at short, reliable delays (inhibition trough: ACx, 10 ms; mPFC, 7 ms), consistent with monosynaptic inhibition generated by inhibitory VIP neurons (Fig. 1d, g, h and Extended Data Fig. 2b). Many, but not all, neurons in this inhibited group had narrow spike widths (ACx, 237 ± 7 µs; mPFC, 225 ± 7 µs; mean $\pm$ s.e.m.) and high firing rates (ACx, 7.4 ± 0.7 Hz; mPFC, 17.5 ± 4.2 Hz; mean $\pm$ s.e.m.; see Extended Data Fig. 2f, g), hallmarks of fast-spiking interneurons, usually expressing parvalbumin (PV)[23]. The inhibited group also contained a subgroup of neurons that was first suppressed by photoactivation of VIP and later activated (Extended Data Fig. 3). A third group of neurons was activated by photostimulation at longer delays and with more temporal spread (Fig. 1e, g, h and Extended Data Fig. 2c). Neurons in this group had wider spikes (ACx, 316 ± 7 µs; mPFC, 339 ± 12 µs; mean $\pm$ s.e.m.) and lower firing rates (ACx, 3.2 ± 0.5 Hz; mPFC, 9.2 ± 2.7 Hz; mean $\pm$ s.e.m.) compared to the inhibited group (Extended Data Fig. 2f, g), indicating that many of these were pyramidal neurons. An analysis of the timing and extent of light-induced firing-rate change revealed that neural responses clustered into three distinguishable groups: short-latency activated followed by inhibited and finally delayed activated neurons (Fig. 1f–h and Extended Data Fig. 2e). This excitation–inhibition–excitation sequence, observed in two functionally different cortical
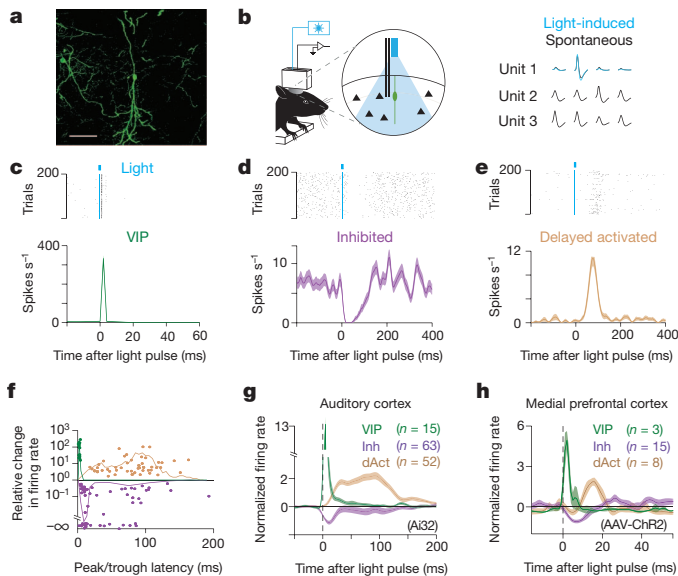
Figure 1 | VIP interneurons generate disinhibition in ACx and mPFC of awake mice. a, Expression of ChR2–YFP in a VIP-Cre mouse. Scale bar, 50 μm. **b**, Left, VIP neurons were identified by optical stimulation *in vivo*. Right, light-evoked spike waveforms (blue) were similar to spontaneous ones (black) (see Methods). **c–e**, Raster plot (top) and peri-stimulus time histogram (PSTH) (bottom) of representative neurons for directly activated (VIP neurons), inhibited and delayed activated groups in ACx. **f**, Relative light-induced firing-rate change (log scale) versus latency of the maximal effect (peak/trough of PSTH). Three separated groups of significantly modulated neurons are apparent: short-latency activated (VIP, green), inhibited (purple) and delayed activated (light brown). Solid lines indicate probability density functions of peak times (normalized separately to improve visibility). **g, h**, Average PSTH of the VIP, inhibited and delayed activated neuron groups in ACx (**g**) and mPFC (**h**). The temporal differences between ACx and mPFC were due to the different ChR2 expression systems (Ai32 versus viral expression, see Extended Data Fig. 9).

regions, is the signature of a disinhibition process: activated VIP interneurons inhibit other interneurons, releasing some pyramidal neurons from inhibitory control. To our knowledge these data represent the first *in vivo* demonstration of cell-type-specific disinhibition, confirming previous suggestions based on connectivity[10–12] that VIP interneurons provide disinhibitory control.

To dissect the circuit mechanisms of disinhibition and identify the cell types that are monosynaptically inhibited by VIP neurons, we turned to the *in vitro* slice preparation. VIP interneurons probably target SOM and/or PV interneurons, the two largest non-overlapping populations (~65% of all interneurons)[19,20]. To examine which of these interneuron subtypes might mediate VIP-initiated disinhibition, we bred VIP-Cre mice with GIN-GFP mice, labelling a subpopulation of SOM-expressing[24] neurons, and separately with G42-GFP mice, labelling a subpopulation of PV-expressing interneurons[25]. Viral delivery of ChR2 into the ACx or mPFC of these mice enabled us to photostimulate VIP interneurons selectively and record the responses of either SOM or PV interneurons identified by epifluorescence *in vitro* (Fig. 2a, b).

Similar to the *in vivo* conditions, photostimulation *in vitro* reliably evoked action potentials in VIP neurons (Extended Data Fig. 4). Activation of VIP interneurons elicited inhibitory postsynaptic currents (IPSCs) from a large fraction of SOM interneurons (Fig. 2a, d). Repeated stimulation of VIP neurons at 40 Hz revealed that these inhibitory connections onto SOM interneurons undergo short-term synaptic depression (Fig. 2a, f). VIP activation elicited IPSCs in a smaller fraction of PV interneurons (Fig. 2b, d). Inhibitory currents (IPSCs) recorded in PV neurons decayed faster than those recorded from SOM neurons (decay time constant: SOM, 18 ± 2 ms; PV, 6 ± 1 ms; mean $\pm$ s.e.m.), and also displayed stronger short-term synaptic depression
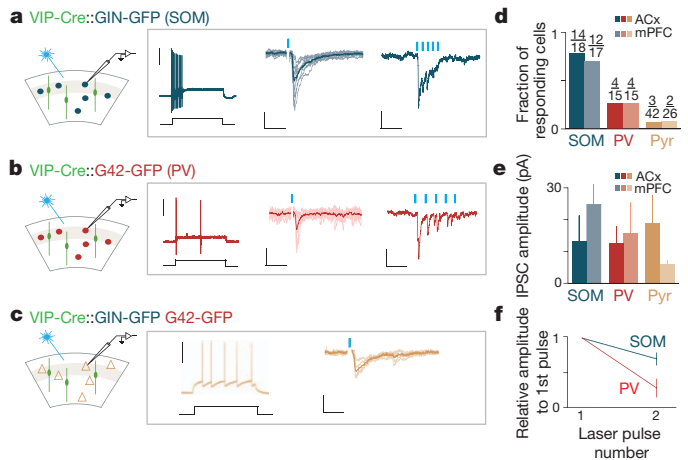
Figure 2 | VIP interneurons inhibit SOM and PV interneurons in ACx and mPFC *in vitro*. a–c, First column: schematic of the *in vitro* experiments. A subpopulation of SOM- (**a**) or PV-expressing (**b**) neurons were identified under epifluorescence. Pyramidal (**c**) cells were identified by soma shape and were sampled from non-fluorescent neurons. Second column: representative firing patterns of SOM, PV and pyramidal neurons during the injection of a depolarizing current (ACx). Scale bar, 40 mV, 50 pA, 200 ms. Third and fourth columns: photostimulation-induced (blue bars) IPSCs from SOM neurons (**a**), PV neurons (**b**) and pyramidal neurons (**c**) at 1 Hz and 40 Hz repetition rates. Scale bars, 1 Hz stimulation: 50 pA, 30 ms (**a**), 40 pA, 25 ms (**b**); and 20 pA, 25 ms (**c**); 40 Hz stimulation: 16 pA, 125 ms (**a**), 75 pA, 50 ms (**b**). **d**, Fraction of neurons responding to photostimulation in ACx and mPFC. **e**, Mean $\pm$ s.e.m. of IPSCs for the significantly responsive neurons (paired *t*-test, $P < 0.01$). Average IPSC amplitudes in ACx and mPFC were not significantly different (*t*-test: SOM, $P = 0.125$; PV, $P = 0.83$; Pyr, $P = 0.256$; note the low sample sizes for pyramidal cells because of the low prevalence of evoked IPSCs). **f**, Short-term depression of IPSCs at 40 Hz (ACx and mPFC combined).

(Fig. 2b, f). In contrast to these inhibitory neurons only a small fraction of pyramidal neurons responded to VIP activation, indicating that pyramidal neurons are a minor monosynaptic target of the VIP population (Fig. 2c, d). The amplitudes of IPSCs were not significantly different across groups (Fig. 2e). These *in vitro* results demonstrate that VIP interneurons specifically inhibit other interneurons, providing a circuit mechanism for a disinhibitory process that we observed *in vivo*.

Because SOM and PV are known to inhibit pyramidal cells, which constitute the majority of cortical neurons, we suspected that the delayed activated neurons *in vivo* were mostly pyramidal neurons. To test this we used immunohistochemistry to map the neural activity marker c-Fos onto identified cell types *in vitro* in mPFC (Extended Data Fig. 5). Photostimulation of VIP neurons increased c-Fos expression fivefold compared to control animals (Extended Data Fig. 5b–i). Eight per cent of c-Fos-immunopositive neurons expressed VIP, whereas the others expressed the pyramidal marker CaMKIIα (Extended Data Fig. 5f, g), revealing that the delayed activated population consists of pyramidal neurons.

To probe the function of this disinhibitory circuit during sensory processing, we investigated how auditory receptive fields are shaped by VIP activation. About one-quarter of the single units recorded in head-restrained awake mice (VIP-Cre::Ai32) (97 of 343) could be classified as directly activated ($n = 4$, VIP, Extended Data Fig. 6), inhibited ($n = 48$), or delayed activated ($n = 46$, Fig. 3a). Delayed activated neurons tended to be more tone responsive (28 of 46 (61%) compared to 97 of 343 (29%) in the entire population) and had stronger auditory responses (Fig. 3b and Extended Data Fig. 7a–f). Also, a large fraction of tone-responsive neurons was delayed activated (28 of 97 (29%), compared to 18 of 245 (7%) in the tone-unresponsive population), and these showed stronger light effects (Fig. 3b, Extended Data Fig. 7a–f and Extended Data Table 1). These results reveal that in auditory
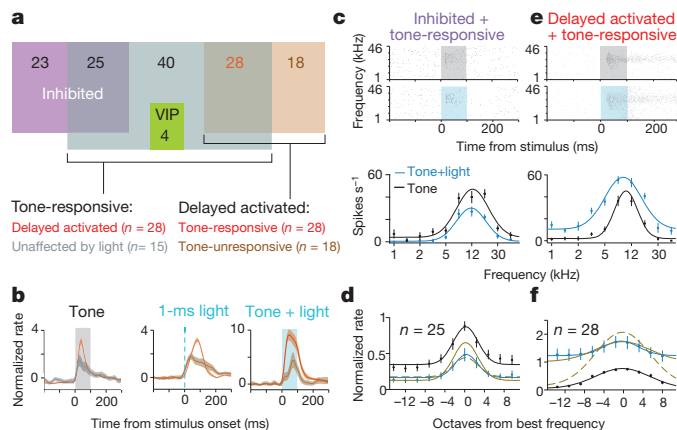
Figure 3 | Auditory responses of a functional subpopulation of principal neurons are modulated by disinhibition. a, Venn diagram showing the number of single neurons in ACx classified on the basis of tone and light responsiveness. **b**, Left: PSTHs aligned to stimulus onset show that tone response of delayed activated cells is stronger than that of those unaffected by light (only cells firing >1 Hz included in the unaffected population, 15/40). Middle, right: light response is stronger in tone-responsive than -unresponsive neurons. **c–f**, Auditory tuning of tone-responsive inhibited (**c, d**) and delayed activated (**e, f**) neurons in awake mice. Mean ± s.e.m. **c, e**, Top: raster plots of representative neurons sorted by the frequency of the auditory stimulus. Grey shade, tone delivery; blue shade, tone plus light stimulation. Bottom: frequency tuning curves of the same cells. **d, f**, Average tuning curves. Mustard solid/dashed lines are predictions from additive/multiplicative models, respectively.

cortex, the VIP circuit disinhibits a functionally specific subset of neurons that tend to be tone responsive and frequency tuned.

Next, we examined how VIP-mediated disinhibition modulates auditory frequency tuning. The inhibited population ($n = 25$ of 97 tone-responsive neurons) decreased whereas delayed activated neurons ($n = 28$ of 97) increased their tone-evoked firing when VIP interneurons were also activated, consistent with the modulation expected from the disinhibitory circuit (Fig. 3c–f). We found that one-parameter gain modulation models (additive or multiplicative) fitted the light-induced changes in average tuning curves: the inhibited population showed divisive gain modulation[9,26], whereas the change in delayed activated neurons was consistent with an additive shift of the baseline firing rate (Fig. 3c–f and Extended Data Fig. 7g). These observations reveal that VIP stimulation modulates the gain of auditory cortical responses.

Having established how VIP activation modulates local circuit activity, next we investigated under what behavioural conditions VIP neurons are recruited. We recorded the activity of VIP neurons during an auditory go/no-go discrimination task (Fig. 4a). VIP neurons were recorded early in training as these sessions had a large number of false-alarm responses to the 'no-go' cue (Fig. 4b). We found that VIP neurons showed surprisingly homogeneous responses to reinforcement feedback signals (Fig. 4c–f and Extended Data Fig. 8). In all VIP neurons, punishment (air puff, $n = 6$, or foot shock, $n = 4$) generated strong phasic activation at short latencies (peak at 50 ± 12 ms, $P < 0.01$). The similar activation in response to two different types of punishment indicates that VIP neurons signal the aversive quality of the negative feedback. Water reward tended to generate weaker and more sustained firing-rate increases (9 of 10 VIP neurons, half-maximum duration 1.2 s; Extended Data Fig. 8). In contrast, unidentified neurons showed heterogeneous responses around the time of reinforcement (Fig. 4e; false-alarm activation, 23 of 130; hit activation, 34 of 130; false-alarm suppression, 25 of 130; hit suppression, 30 of 130 or no change). A small subpopulation of unidentified neurons responded to both reward and punishment (17 of 130). However, these responses tended to precede the feedback, indicating sustained auditory responses to the cue, unlike the abrupt firing-rate
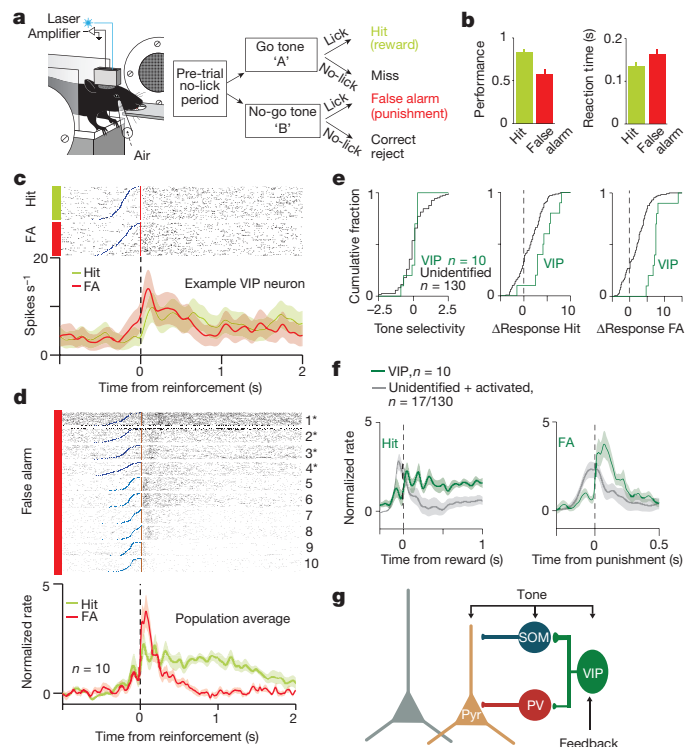
Figure 4 | VIP neurons are recruited by reinforcement signals. a, Schematic of behavioural setup and auditory discrimination task. **b**, Average performance and reaction time (mean ± s.e.m. across sessions). **c**, Raster plots and peri-event time histograms (PETH) of example VIP neuron. Neural activity was aligned to reinforcement (reward/punishment; red). Blue, tone onset. FA, false alarm. **d**, Raster plots and average PETHs of VIP neurons ($n = 10$). All VIP neurons showed strong increase of firing rate after punishment (VIP 1–4, foot shock (asterisk); 5–10, air puff). **e**, Cumulative fraction of tone selectivity (left; Kolmogorov–Smirnov test, $P = 0.43$) and firing-rate change (right; hit, $P < 0.01$; false alarm, $P < 0.01$; ΔResponse, see Methods). Response to reinforcement signals distinguishes VIP neurons from unidentified population. **f**, Normalized average PETH aligned to reinforcement of VIP and concurrently recorded unidentified neurons responsive to both positive and negative feedback (17 of 130; grey). Left: VIP neurons showed sustained activation after reward delivery. Right: VIP neurons increased firing rate abruptly after punishment, in contrast with unidentified cells. **g**, Schematic model of disinhibitory circuit. Feedback information (for example, reinforcement signals) to VIP neurons disinhibits a functional subpopulation of pyramidal neurons.

increase of VIP interneurons after punishment delivery (Fig. 4f and Extended Data Fig. 8c). These results reveal that VIP interneurons respond in unison to reinforcement feedback, distinct from the more diverse responses observed in the rest of the cortical population.

Genetic targeting and optical activation enabled us to map a particular circuit function, disinhibitory control, onto a molecularly defined cell type. On the basis of the strong disinhibitory impact and target selectivity observed in two cytoarchitectonically and functionally different cortical areas, we propose that the disinhibitory microcircuit mediated by VIP-expressing interneurons represents a conserved motif in neocortex.

VIP interneurons function within a highly interconnected network, therefore their role can be understood as pre-synaptic drivers of ('impact') and post-synaptic responders to ('recruitment') other neurons. In terms of impact, we found that VIP neurons mediate disinhibitory control. In terms of recruitment, we identified a behaviourally relevant condition, reinforcement feedback, that uniformly activates VIP neurons. The homogeneous behavioural recruitment of VIP interneurons indicates that the synchronous ChR2-mediated activation to probe their circuit function was physiologically plausible. VIP interneurons are ideally positioned to serve as a substrate for long-range

inputs to increase the gain of local cortical processing (Fig. 4g). Studies have recently demonstrated a similar disinhibitory process[13,18], whereby foot-shock-induced cholinergic activation of layer 1 cortical neurons in ACx enabled auditory fear learning[18]. Interestingly, most VIP interneurons are located in superficial layers, including layer 1[19,20] (Extended Data Fig. 1g, h). However, the extent to which these two disinhibitory circuits overlap remains to be determined, as only a fraction of layer 1 interneurons express VIP. VIP interneurons express ionotropic receptors for cholinergic (nAChR) and serotonergic (5HT3a) modulation[27–29], indicating that they are also subject to rapid neuromodulation[30]. These neuromodulatory systems or other long-range pathways probably convey information about reinforcement events to VIP neurons. By rapidly relaying this signal primarily to tone-selective neurons locally, VIP interneurons might contribute to cortical learning mechanisms. On the basis of these observations we propose that disinhibitory control by VIP interneurons provides a powerful circuit mechanism that enables long-range cortical signals or subcortical neuromodulation to efficiently modulate specific pyramidal neuron ensembles.

METHODS SUMMARY

All animal procedures were performed in accordance with National Institutes of Health standards and were approved by Cold Spring Harbour Laboratory Institutional Animal Care and Use Committee.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Silberberg, G. & Markram, H. Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. *Neuron* **53,** 735–746 (2007).
2. Kapfer, C., Glickfeld, L. L., Atallah, B. V. & Scanziani, M. Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex. *Nature Neurosci.* **10,** 743–753 (2007).
3. Gentet, L. J. et al. Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. *Nature Neurosci.* **15,** 607–612 (2012).
4. Cardin, J. A. et al. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* **459,** 663–667 (2009).
5. Sohal, V. S., Zhang, F., Yizhar, O. & Deisseroth, K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459,** 698–702 (2009).
6. Isaacson, J. S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72,** 231–243 (2011).
7. Ascoli, G. A. et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Rev. Neurosci.* **9,** 557–568 (2008).
8. Carandini, M., Heeger, D. J. & Movshon, J. A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17,** 8621–8644 (1997).
9. Salinas, E. & Thier, P. Gain modulation: a major computational principle of the central nervous system. *Neuron* **27,** 15–21 (2000).
10. Acsády, L., Gorcs, T. J. & Freund, T. F. Different populations of vasoactive intestinal polypeptide-immunoreactive interneurons are specialized to control pyramidal cells or interneurons in the hippocampus. *Neuroscience* **73,** 317–334 (1996).
11. Dávid, C., Schleicher, A., Zuschratter, W. & Staiger, J. F. The innervation of parvalbumin-containing interneurons by VIP-immunopositive interneurons in the primary somatosensory cortex of the adult rat. *Eur. J. Neurosci.* **25,** 2329–2340 (2007).
12. Hájos, N., Acsady, L. & Freund, T. F. Target selectivity and neurochemical characteristics of VIP-immunoreactive interneurons in the rat dentate gyrus. *Eur. J. Neurosci.* **8,** 1415–1431 (1996).
13. Jiang, X., Wang, G., Lee, A. J., Stornetta, R. L. & Zhu, J. J. The organization of two new cortical interneuronal circuits. *Nature Neurosci.* **16,** 210–218 (2013).
14. Xu, H., Jeong, H. Y., Tremblay, R. & Rudy, B. Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4. *Neuron* **77,** 155–167 (2013).
15. Pfeffer, C. K., Xue, M., He, M., Huang, Z. J. & Scanziani, M. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature Neurosci.* **16,** 1068–1076 (2013).
16. Lovett-Barron, M. et al. Regulation of neuronal input transformations by tunable dendritic inhibition. *Nature Neurosci.* **15,** 423–430 (2012).
17. Lee, S. H. et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488,** 379–383 (2012).
18. Letzkus, J. J. et al. A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* **480,** 331–335 (2011).
19. Xu, X., Roby, K. D. & Callaway, E. M. Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *J. Comp. Neurol.* **518,** 389–404 (2010).
20. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71,** 45–61 (2011).
21. Taniguchi, H. et al. A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71,** 995–1013 (2011).
22. Madisen, L. et al. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neurosci.* **15,** 793–802 (2012).
23. Csicsvari, J., Hirase, H., Czurko, A. & Buzsaki, G. Reliability and state dependence of pyramidal cell-interneuron synapses in the behaving rat: an ensemble approach in the behaving rat. *Neuron* **21,** 179–189 (1998).
24. Oliva, A. A. Jr, Jiang, M., Lam, T., Smith, K. L. & Swann, J. W. Novel hippocampal interneuronal subtypes identified using transgenic mice that express green fluorescent protein in GABAergic interneurons. *J. Neurosci.* **20,** 3354–3368 (2000).
25. Chattopadhyaya, B. et al. Experience and activity-dependent maturation of perisomatic GABAergic innervation in primary visual cortex during a postnatal critical period. *J. Neurosci.* **24,** 9598–9611 (2004).
26. Williford, T. & Maunsell, J. H. Effects of spatial attention on contrast response functions in macaque area V4. *J. Neurophysiol.* **96,** 40–54 (2006).
27. Paspalas, C. D. & Papadopoulos, G. C. Serotoninergic afferents preferentially innervate distinct subclasses of peptidergic interneurons in the rat visual cortex. *Brain Res.* **891,** 158–167 (2001).
28. Kawaguchi, Y. Selective cholinergic modulation of cortical GABAergic cell subtypes. *J. Neurophysiol.* **78,** 1743–1747 (1997).
29. Arroyo, S., Bennett, C., Aziz, D., Brown, S. P. & Hestrin, S. Prolonged disynaptic inhibition in the cortex mediated by slow, non-alpha7 nicotinic excitation of a specific subset of cortical interneurons. *J. Neurosci.* **32,** 3859–3864 (2012).
30. Alitto, H. J. & Dan, Y. Cell-type-specific modulation of neocortical activity by basal forebrain input. *Front Syst Neurosci* **6,** 79 (2013).

Author Contributions H.-J.P., B.H. and A.K. designed the experiments. H.-J.P. and B.H. performed the experiments and analysed data. D.K. set up *in vivo* optogenetics-assisted recordings. J.I.S. designed custom behaviour and stimulation systems for the behavioural task. Z.J.H. provided the VIP-IRES-Cre mouse line. H.-J.P., B.H. and A.K. wrote the manuscript with comments from Z.J.H., J.S. and D.K.

METHODS

Animals. Adult (over 2 months old) male and female mice of VIP-Cre[21] (C57BL/6 background) or VIP-Cre crossed with Ai32 (ref. 22) (ChR2 reporter line), GIN-GFP (SOM) and G42-GFP (PV) were used under the protocol approved by Cold Spring Harbour Laboratory Institutional Animal Care and Use Committee in accordance with National Institutes of Health regulations.

Virus injection. Animals were anaesthetized with ketamine (100 mg kg^{-1}) and xylazine (10 mg kg^{-1}). AAV2/9.EF1α.DIO.ChETA.EYFP (ref. 31) (UNC vector core) was injected in mPFC (1.75 mm anterior to bregma and 0.5 mm lateral to midline) or ACx (2.50 mm posterior to bregma and 4.00 mm lateral to midline) of VIP-Cre mice, VIP-Cre::GIN-GFP and VIP-Cre::G42-GFP at 4 weeks of age. Approximately 1 μl of AAV (8×10^{12} virus particles per ml) was injected with a glass pipette using Picospritzer (Parker Hannifin Co.). Because VIP cells constitute only a small fraction (1–2%) of cortical neurons, the expression of ChR2 was maintained at least for 4 weeks. The delivery of large volumes of AAV at 4 weeks of age and the long expression time increased the efficacy and reduced the variability of ChR2 expression. The longer expression time (even after 6 month) did not affect animals' health and behaviour.

Neural data collection. After 4–6 weeks of ChR2 expression, a custom-built drive housing 6–8 tetrodes and 1–2 optical fibres (50 μm core diameter, numerical aperture 0.2, Polymicro Technologies) were implanted in the left mPFC (1.75 mm anterior to bregma and 0.5 mm lateral to midline) or the left ACx (2.50 mm posterior to bregma and 4.00 mm lateral to midline) using stereotaxis[32]. For frequency tuning and auditory go/no-go experiments, a titanium headbar was also attached to the skull. After 7–10 days of recovery after surgery, action potentials were recorded extracellularly (sampled at 32 kHz) with either a Cheetah32 or a DigiLynx system (Neuralynx, Inc.) from the ACx ($n = 12$ mice) or the mPFC ($n = 4$ mice). Brief laser pulses (473 nm, 1-ms duration, 1.5–3 mW of total output at the tip of optical fibres) were delivered through an optical fibre (50 μm core diameter, numerical aperture 0.2, CrystaLaser, UltraLasers or Lasermate Group Inc.). Electrodes were advanced 60–80 μm each recording day. Electrode locations were estimated based on the entry coordinates and the extent of cumulative descent and later confirmed by histology (Extended Data Fig. 1e, f). In most of the ACx experiments (50 of 53 sessions, $n = 12$ mice), frequency tuning experiments were performed after optogenetic tagging while neural signals were continuously recorded. Therefore, the number of neurons in Figs 3 and 4 are a subset of Fig. 1. Conducting the experiments required monitoring the effect of VIP-photostimulation on multiunit activity; therefore, it was not possible to fully blind the experimenter. However, this information was only used to guide decisions about whether to move the electrodes or not. Once light effects were detected, all well-separated single units from the session were analysed by automated software algorithms treating every neuron equally. Sample size was estimated on the basis of previous literature[18].

The difference in the relative proportion of recorded VIP neurons in ACx and mPFC is probably due to technical reasons. First, ChR2 was delivered virally into mPFC, whereas we used a reporter line (Ai32) for ACx. Therefore, ChR2 expression in mPFC was spatially limited and the level of expression could vary. Second, superficial layers in mPFC were more difficult to target because they are located underneath midline blood vessels. There was no significant difference in VIP cell density in ACx and mPFC. Cell densities: ACx, 2.1 ± 0.17 cells 10^{-4} μm^{-2} (115 cells, 4 slices from 2 mice); mPFC, 2.1 ± 0.09 cells 10^{-4} μm^{-2} (124 cells, 4 slices 2 mice); $P = 0.78$, t-test. Cell density was estimated from VIP.Ai32 mice.

Data analysis. All data analysis was carried out using built-in and custom-built software in Matlab (Mathworks). All recording sessions with light effects were included. Spikes were manually sorted into clusters (presumptive neurons) off-line based on peak amplitude and waveform energy using the MClust software (A. D. Redish). Cluster quality was quantified using isolation distance and L-ratio[33]. Putative cells with isolation distance <20 or L-ratio >0.1 were excluded. Autocorrelation functions were inspected for all putative cells and in cases with absolute refractory period violations, an additional effort was made to improve cluster separation. If refractory violations persisted, the cluster was excluded. Some of the VIP neurons did not reach the threshold of isolation distance and L-ratio. For these neurons, we exploited the waveform information carried by light-evoked action potentials, which aided the isolation of these units. We calculated waveform correlation between spontaneous action potentials of VIP neurons and average light-evoked waveform. We restricted the clusters to the upper 5 percentile of the bootstrapped distribution of correlation coefficients.

Next, we estimated peri-stimulus (stimulus-aligned) firing rates by using an adaptive spike density function (SDF) approach (termed peri-stimulus time histogram (PSTH)). Briefly, spike rasters were convolved with a variable kernel Gaussian window to provide a SDF estimate. The kernel width of the Gaussian was adapted to the local estimate of spiking probability to implement stronger smoothing when information was sparse. Variance was mapped onto spiking

probability between 0 (moving average, corresponding to probability of 0) and infinity (Dirac-delta, corresponding to probability of 1). To detect light-induced changes of firing rate we first determined the putative activation/suppression period and then evaluated the statistical significance of the firing rate change compared to a stimulus-free baseline, as follows. Adaptive SDF was calculated aligned to light stimulus onset. For the mPFC, minimal and maximal firing was determined as the minimum and maximum of the SDF within 100 ms from the light pulses. The baseline firing rate was calculated from mean firing probability within a 100-ms window before the start of a pulse train. For ACx, longer windows were used according to the observed differences in response latencies (200 ms after light flashes for response and 400 ms before light pulses for baseline). We determined the putative activation period as the epoch between the half-peak crossings (relative to baseline) before and after the positive peak. Putative suppression period was defined in a similar way based on the negative peak. The statistical significance of the activation and suppression was determined by comparing the spike count distribution within these periods with an equivalent baseline epoch using a two-tailed Mann–Whitney U-test. A P value cutoff of 0.05 was used for significance testing. In the mPFC experiments, light pulse trains of different frequencies were used as stimuli. To allow a long enough window for testing possible firing-rate changes, single pulses of the slowest stimulus bursts (5 Hz) were chosen as reference events for this analysis. Because slow light-induced firing-rate changes could potentially mask fast light-evoked activation or suppression, the analysis was repeated, restricted to first stimuli of light pulse trains; a neuron was considered activated or inhibited if either of the two tests showed a significant effect. In ACx experiments, single pulses of 0.5 Hz frequency were applied; this slow stimulation protocol allowed us to include all pulses as reference events, enhancing statistical power. To reduce the probability of misclassification of ACx neurons based on light-induced firing-rate changes to a minimum, we exploited tone plus light stimulation trials for those cells showing any effects for 1-ms laser light stimulation. If any of these cells showed additional effects when including the tone plus light stimuli without confounding effects of tone only stimulation, then these effects were also taken into account. Significance of suppression could not reliably be determined for cells with a baseline firing rate <2 Hz (<1 Hz in ACx; relaxing the spike rate threshold was enabled by the improved statistical power, see above). In a few cases we detected late secondary effects after a significant short latency inhibition or activation. These cells were classified based on the primary effects. In the ACx experiments, a significant portion of the neurons that were inhibited also showed a later activation. These cells were also included in the inhibited group in the main figures based on their significant short-latency inhibition after light stimulation. In addition, they are also displayed separately in Extended Data Fig. 3.

The timing of firing-rate suppression in neurons significantly inhibited after photostimulation was compared for narrow spiking (<275 μs, $n = 17$) and wide spiking (>275 μs, $n = 46$) neurons. Wide spiking neurons showed significantly later offset (mean ± s.e.m., narrow spiking, 39.0 ± 10.0 ms; wide spiking, 86.5 ± 7.6 ms; $P = 0.0007$, Mann–Whitney U-test for all comparisons in this paragraph) and correspondingly longer duration (narrow spiking, 27.3 ± 5.8 ms; wide spiking, 54.3 ± 5.8 ms, $P = 0.0019$) of suppression. Similar results were obtained when comparing slow firing (<5 Hz, $n = 32$) and fast spiking (>5 Hz, $n = 31$) neurons: slow firing neurons showed significantly later inhibition offset (fast spiking, 58.6 ± 9.2 ms; slow firing, 88.4 ± 9.1 ms; $P = 0.016$) and longer durations (fast spiking, 34.6 ± 4.8 ms; slow firing, 59.0 ± 7.6 ms; $P = 0.0051$).

***In vitro* electrophysiology.** For *in vitro* electrophysiology experiments, ChR2 expression time was kept constant and comparable (4–6 weeks) between VIP-Cre::GIN-GFP and VIP-Cre::G42-GFP. Brain slices were prepared at 8–10 weeks of age. Mice were anaesthetized and decapitated. The brain was transferred to an ice-cold cutting solution containing (in mM) 110 choline chloride, 25 NaHCO$_3$, 25 D-glucose, 11.6 sodium ascorbate, 3.1 sodium pyruvate, 2.5 KCl, 1.25 NaH$_2$PO$_4$, 7 MgCl$_2$ and 0.5 CaCl$_2$. Coronal sections (300 μm) of mPFC or ACx were prepared using vibratome (Microm) and all slices were transferred to artificial cerebrospinal fluid (ACSF) containing (in mM) 127 NaCl, 25 NaHCO$_3$, 25 D-glucose, 2.5 KCl, 1.25 NaH$_2$PO$_4$, 2CaCl$_2$ and 1MgCl$_2$, balanced with 95% O$_2$ and 5% CO$_2$. Slices were incubated at 34 °C for 30–60 min and kept at room temperature (22–24 °C) during the experiments. Target neurons were identified by fluorescence and patched with glass electrodes (pulled from borosilicate glass capillaries, Warner Instruments, resistance, 5–7 MΩ). To augment IPSCs in inward direction, a high Cl$^-$ intracellular solution containing (in mM) 30 potassium-methlysulphate, 118 KCl, 4 MgCl$_2$, 10 HEPES, 1 EGTA, 4 Na$_2$-ATP, 0.4 Na2-GTP, 10 sodium phosphocreatine, pH 7.25; 290–300 mOsm was used. To block excitatory postsynaptic currents 20 μM CNQX (Tocris) and 50 μM APV (Tocris) were added to the ACSF. Whole-cell recordings were conducted and signals were amplified using Multiclamp 700A amplifier (Axon Instruments, Molecular Devices). IPSCs were measured in voltage clamp mode at a holding potential of −70 mV, and action

potentials were recorded in current clamp mode. VIP neurons expressing ChR2(ChETA) were activated by 1-ms light pulses at 1 Hz or 40 Hz and only fired single action potentials in response to single light pulses. All recording sessions with light effects were included and analysed using a custom-built Matlab program (Mathworks). Neurons that were significantly responsive to photostimulation (peak amplitude versus baseline, paired t-test, $P < 0.01$) were further analysed to calculate the IPSC amplitude. The normality of amplitude distribution was tested on a subset of data with the Lilliefors-test. Decay time constants were calculated by fitting the decaying phase of the IPSCs (10–90% of the peak amplitude) with a single exponential function. There were no significant differences in IPSC amplitude and fraction of responded neurons in different cortical layers and the results were combined. SOM and PV showed similar short-term plasticity in ACx and mPFC. Hence, the two data sets were combined. All central tendencies were reported as mean ± s.e.m.

Frequency tuning experiment and data analysis. Experiments were performed in a sound attenuated chamber (Industrial Acoustics Co.). Animals were head-restrained using headbars and monitored during recording sessions using a USB camera. Before starting recording, the mice were accommodated to the head-fixing setup. Untrained awake mice sometimes moved in bouts in head-restrained condition. To examine whether these movement bouts modulate neural activity in ACx, a 3.8×3.8 cm force-sensitive resistor (Interlink electronics) placed under the paws was used to monitor movement while neural activity was recorded simultaneously. Neural activity was sorted according to epochs of movement and quiescence and the average firing rates between the two conditions were compared statistically (unpaired t-test; $P < 0.01$). The analysis revealed that most neurons (40 of 43) did not show a statistically significant difference, suggesting that the movement in head-restrained condition was not reflected in substantial changes of neural activity in the ACx. Pure tones (1–46 kHz, 100 ms, 50–70 dB) were delivered in a pseudo-random sequence to the right ear (TDT) with or without concurrent photostimulation (100-ms duration (except one session with 200 ms), 473 nm, 1.5–3 mW of total output at the tip of the optical fibres; 50 μm core diameter, numerical aperture 0.2, Polymicro Technologies, Ultralasers) in awake state. Single-neuron activity was recorded with Cheetah system (Neuralynx). Sound presentation and neural data were synchronized by acquiring time-stamps from the sound-delivery system along with the electrophysiological signals.

Data were analysed using Matlab (Mathworks). All central tendencies were reported as mean ± s.e.m. To classify tone-responsive neurons, adaptive spike density function (SDF) was calculated by convolving the raster plots of tone responses with a variable Gaussian window (see above). Minimal/maximal firing was assessed as minimum/maximum SDF within 200 ms from tone onset. Baseline firing was determined by mean pre-event firing probability. The time course of the firing rate change was assessed by crossings of the half-distance between the extreme and the baseline before and after the minimum/maximum. This temporal window of increase/decrease was then used to bin the baseline raster. Spike counts for baseline and spike counts in the previously determined temporal window were compared using Mann–Whitney U-test and the cells with P value <0.01 were classified as tone-responsive neurons. Firing rates during pure tone or tone plus light stimuli (100 ms) were calculated to obtain auditory tuning curves. A subgroup of inhibited cells showed secondary, delayed activation by VIP photostimulation after the inhibition ceased (Extended Data Fig. 3). For these cells, only the first 25 ms of the stimuli were used for the tuning curves to avoid any confounding effects resulting from the delayed activation. These cells were also analysed separately in more detail using two different stimulus-induced firing rate windows (0–25 ms for inhibition and 75–100 ms for delayed activation; see Extended Data Fig. 3). For the tone-responsive neurons not modulated by 1-ms light pulses, only those cells with a baseline firing rate >1 Hz (15 of 40) were analysed to ensure reliable exclusion of potential inhibition after photostimulation (see above). Tuning curves of tone-responsive neurons were fit by Gaussian curves (robust fit with nonlinear least squares method) and aligned for averaging based on the mean of the Gaussians. The underlying raw firing rates for both tone only and tone plus light stimulation protocols were normalized to the maximum firing rate in the tone only stimulation condition and averaged over cells for different groups. The averages were also fit by Gaussian curves.

We quantified the effects of VIP stimulation on auditory tuning curves by fitting simple models. First, we considered two one-parameter models: pure divisive gain (activity gain)[26] and a mere shift in baseline firing rate. Although the individual data showed considerable heterogeneity, on average the decrease in firing rate in the inhibited neurons was consistent with an activity gain model, whereas the increased rates in the delayed activated cells were best fit with a baseline shift (Fig. 3 and Extended Data Fig. 7g). However, nonlinear transformations such as a threshold operation in combination with a baseline shift can in theory yield similar or better results compared to the activity gain model[26]. We tested this combined

model for inhibited cells and compared it with the results of the activity gain model. The activity gain model resulted in better fits (smaller least square errors) than the baseline shift combined with a threshold operation at 0 Hz in the case of 22 out of 25 neurons; the combined model resulted in smaller errors in 2 neurons and the fits were indistinguishable for 1 cell. This suggests that the activity gain model is superior to the additive model for inhibited cells even when allowing an additional nonlinear modulation ($P = 2.7 \times 10^{-5}$, Wilcoxon signed rank test, two-tailed).

Next, we turned to two-parameter models. Because the best frequency and the tuning width did not show substantial changes (inhibited group, change in best frequency, 0.86 kHz; tuning width quantified by the s.d. of the Gaussian fit, 3.12 (95% confidence intervals, 2.68, 3.56) without and 3.51 (95% confidence intervals, 2.62, 4.41) with VIP stimulation; activated group, change in best frequency, 2.64 kHz; tuning width quantified by the s.d. of the Gaussian fit, 6.95 (95% confidence intervals, 6.25, 7.65) without and 5.37 (95% confidence intervals, 4.31, 6.42) with VIP stimulation), we tested two-parameter models combining the baseline shift and the activity gain model. This model resulted in only marginal improvement relative to the divisive model for the inhibited group (Bayesian information criterion[34] for the combined, divisive gain and baseline shift models: −399.6, −398.3 and −354.7) and the baseline shift model for the delayed activated group (Bayesian information criterion for the combined, multiplicative gain and baseline shift models: −274.7, −146.7 and −270.2). Thus, the effect of VIP stimulation on auditory tuning curves is explained by the divisive gain modulation of inhibited cells and the additive firing-rate change of the delayed activated neurons.

Auditory go/no-go task. Mice were trained on a head-restrained auditory discrimination task. The animals were placed in a head-restrained setup inside a sound-attenuated chamber (Industrial Acoustics Co.) monitored by a USB video camera (LE). Mice were trained to lick for water reward (3–5 μl per trial) after hearing a go tone (5 or 20 kHz frequency, 0.5 s duration) while withholding response to a no-go tone (10 or 4 kHz frequency, 0.5 s duration) associated with punishment (gentle air puff, 100 ms duration, $n = 5$ mice or mild foot shock, 100 ms duration, 0.6 mA, $n = 1$ mouse). Licks were detected by an infrared lick-ometer (Island Motion Co.). Laser pulse sequences and tones were generated using Pulse Pal, a custom 4-channel stimulator we developed based on an open source ARM microcontroller platform (Maple, LeafLabs). All behavioural events were acquired using Bpod, a custom-designed behavioural system that provides a real-time virtual state machine framework for precise control of stimulus delivery and environmental measurements. Events were synchronized with neural signals by digital inputs from Bpod to the recording system (Neuralynx). Each trial was concluded with one of the four following possible outcomes: two types of correct trials, hit (response to go tone) or correct rejection (no response to no-go tone) and two types of incorrect trials, miss (no response to go tone) or false alarm (response to no-go tone) (Fig. 4a). Tones were delivered in a pseudo-random sequence. Once mice stopped licking, the next trial started with a pseudo-random delay (average 5 s or 1.75 s). The animals were trained to perform 200–300 trials to allow firing rate comparisons in >100 ms windows. However, individual sessions varied between 96–600 trials.

Peri-event time histograms (PETH) aligned to task events (cue tone onset, reinforcement signals) were calculated using the adaptive SDF algorithm described above. Assessment of significant firing-rate changes after task events was performed similarly to the analysis of photostimulation- or tone-evoked enhancement or suppression of activity (see above). Neurons were grouped according to their firing-rate responses around reinforcement signals (baseline window, −1 to −0.6 s before cue tone onset; test window, 0 to 0.4 s after cue tone onset). PETHs were z-scored, averaged and baseline-subtracted within groups. Tone selectivity and ΔResponses of hit and false alarm were calculated from maximum/minimum values of normalized PETHs during go/no-go tone periods. Tone selectivity was defined as |go tone response − no-go tone response|/|go tone response + no-go tone response|. ΔResponse was defined as maximum/minimum rate − baseline rate.

c-Fos experiments and quantification. A single optical fibre (50 μm core diameter, numerical aperture 0.2, Polymicro Technologies) was implanted in the left mPFC of VIP-Cre mice with or without ChR2 expression (4–6 weeks). After 7–10 days of recovery post surgery, animals were anaesthetized with ketamine (100 mg kg^{-1}) and xylazine (10 mg kg^{-1}) and left in their home cages in a dark, sound attenuated chamber for 2–3 h to reduce background c-Fos levels. The photostimulation protocol was as follows: stimulation was composed of a train of 20 brief (1 ms) laser pulses for 1 s followed by no stimulation for 3 s. This was repeated 25 times (total stimulation time = 100 s). Animals were put back to their home cages in the sound-attenuated chamber for 1 h. Animals were then perfused transcardially with a fixative containing 4% paraformaldehyde in 0.1 M phosphate buffer and the brains were postfixed overnight in the same solution at 4 °C. After sectioning and confocal imaging (see below), the z-stacked images were analysed

for quantification of c-Fos and overlap with other markers (CaMKIIα, ChR2-YFP). The quantification was performed by manual counting of nuclear staining.

Immunofluorescence and imaging. Animals were perfused according to the same protocol used for c-Fos experiments (see above). Brain sections (80–120 μm thickness) were prepared using a vibratome (Leika), and permeabilized and blocked with 1% Triton X-100 and 5% normal goat serum for 2 h. Immunostaining was performed with primary antibodies of rabbit anti-VIP (ImmunoStar, 1:400 dilution); rabbit anti-c-Fos (Santa Cruz Biotechnology, 1:1,000); mouse anti-CaMKIIα (Thermo Scientific, 1:200); mouse anti-NeuN (Chemicon International Inc., 1:500) in 0.1% Triton X-100 and 5% normal goat serum in PBS overnight at 4 °C. After three washes of 5 min in PBS, sections were incubated with secondary antibodies (Alexa-596/Alexa-405 conjugated goat anti-rabbit/anti-mouse, Invitrogen/Molecular Probes) diluted 1:500 in 0.1% Triton X-100 and 5% normal goat serum in PBS. Z-stack images were taken using a confocal microscope (Zeiss 710 LSM).

Immunostaining was used to estimate specificity and efficiency of VIP expression in the VIP-Cre:Ai32 line. However, we would like to note that this approach has some limitations, as immunohistochemistry is not considered a gold standard method for assessing gene expression, as both its specificity and sensitivity depends on external factors[21,35]. These factors include but are not restricted to (1) cell-type-specific availability of the epitope[36]; (2) varying expression levels and subcellular patterns of the antigens[35,37]; (3) details of the experimental procedure[35].

31. Zhang, F. et al. The microbial opsin family of optogenetic tools. *Cell* **147,** 1446–1457 (2011).
32. Kvitsiani, D. et al. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498,** 363–366 (2013).
33. Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K. & Redish, A. D. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience* **131,** 1–11 (2005).
34. Schwarz, G. E. Estimating the dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).
35. Lorincz, A. & Nusser, Z. Cell-type-dependent molecular composition of the axon initial segment. *J. Neurosci.* **28,** 14329–14340 (2008).
36. Katona, I. et al. Molecular composition of the endocannabinoid system at glutamatergic synapses. *J. Neurosci.* **26,** 5628–5637 (2006).
37. Varga, V. et al. The presence of pacemaker HCN channels identifies theta rhythmic GABAergic neurons in the medial septum. *J. Physiol.* **586,** 3893–3915 (2008).
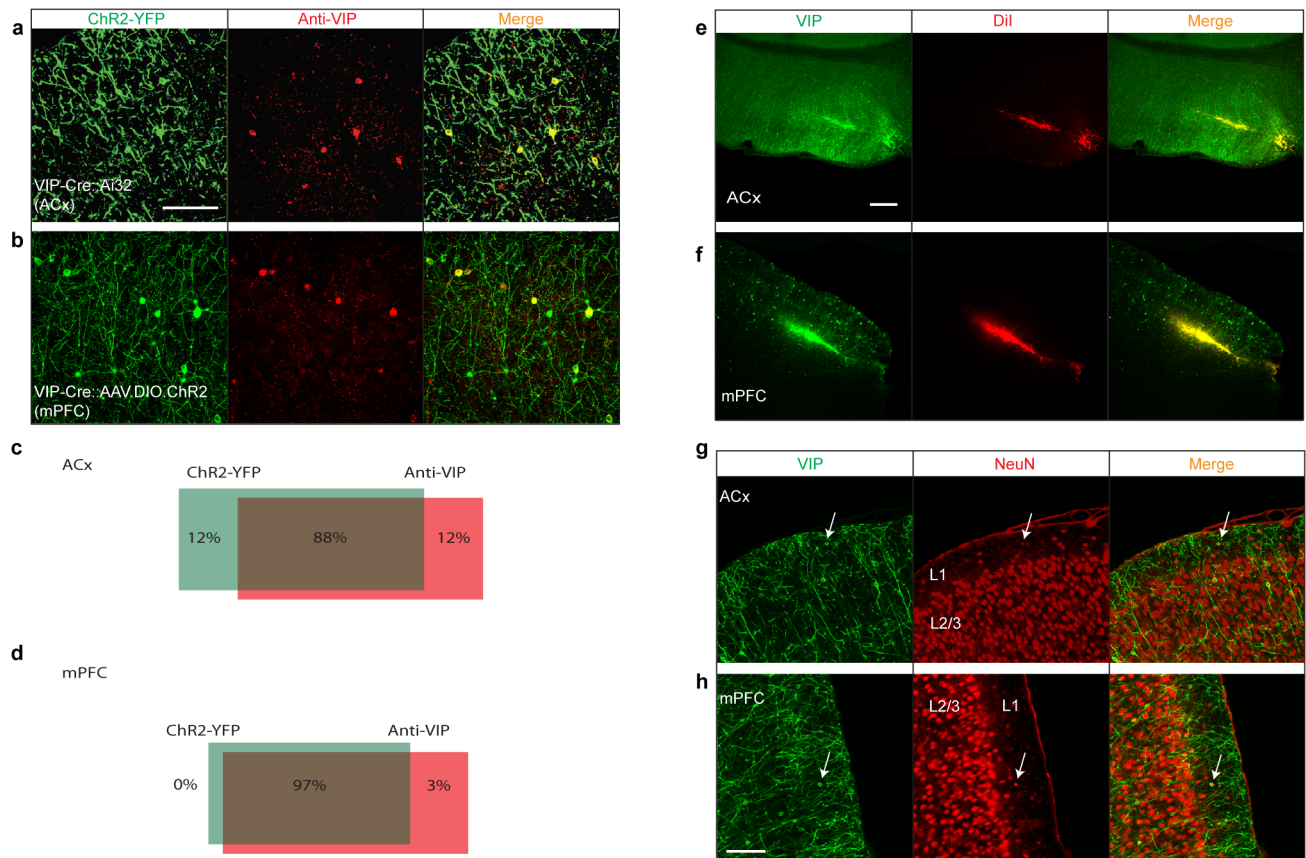
Extended Data Figure 1 | Specificity and efficiency of ChR2 expression, recording locations and layer 1 VIP neurons. **a**, VIP-Cre:Ai32 (ACx). **b**, VIP-Cre::AAV-DIO-ChR2 (mPFC). Note that some green puncta were not somata and only somata were used for quantification. **c**, Quantification of the overlap of ChR2–YFP (green) with anti-VIP (red) in ACx. Overlap = 88 ± 6.6% (49 of 54 neurons), 6 slices from 4 mice (see Methods for a note on caveats). **d**, Quantification of the overlap of ChR2–YFP (green) with anti-VIP (red) in mPFC. Overlap = 97 ± 3.7% (35 of 36 neurons), 5 slices from 4 mice. Scale bar, 50 μm. **e**, **f**, Recording location in mPFC and ACx. Microdrives accommodating 6 tetrodes and 1 optical fibre were implanted in the ACx (**e**) or the mPFC (**f**). Recording sites were confirmed by histology using DiI (red) that was applied to the optical fibre before implantation. Histology results showed that the electrode locations were biased towards the middle layers. Green, VIP neurons expressing ChR2–YFP; red, DiI. Scale bar, 200 μm. **g**, **h**, Most VIP neurons were located in layer 2/3, with a smaller fraction in layer 1. VIP comprised about 10% of layer 1 neurons. **g**, VIP neuron in layer 1 of the ACx (arrow; 6 VIP/60 layer 1 ACx neurons, $n = 6$ slices from 2 mice). **h**, VIP neuron in layer 1 of the mPFC (5 VIP/56 layer 1 mPFC neurons, $n = 6$ slices from 2 mice). Green, VIP; red, NeuN (neuronal marker) staining. Scale bar, 100 μm.
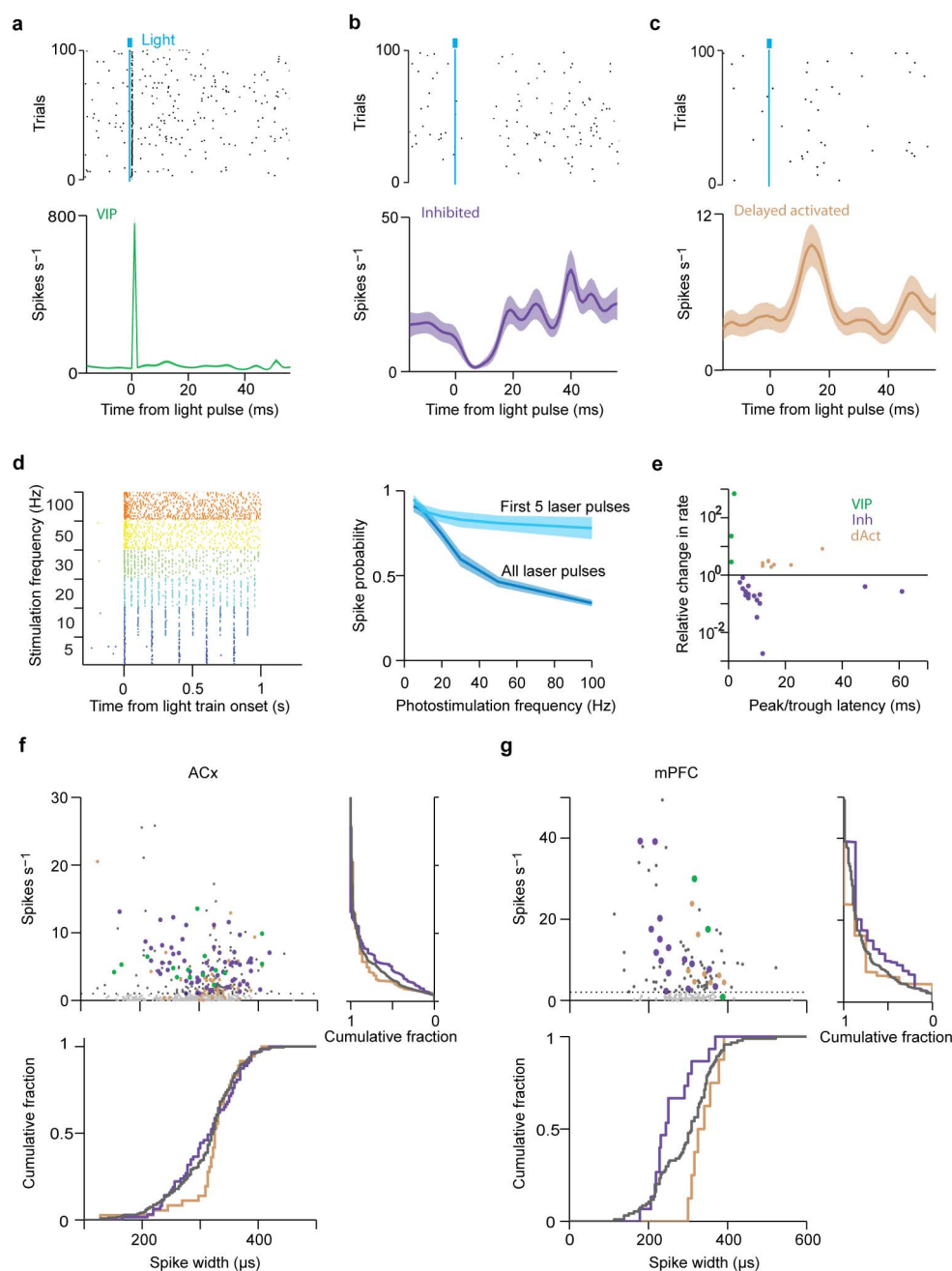
Extended Data Figure 2 | Three distinct populations responsive to photostimulation in mPFC and spike width versus firing rate. a–c, Raster plots and PSTHs aligned to photostimulation for three distinct populations in mPFC. Examples of a directly activated (VIP, **a**), an inhibited (**b**) and a delayed activated neuron (**c**). Stimulation frequency, 10–20 Hz. **d**, Photostimulation-evoked spike probability of a VIP interneuron. Left: raster plot. Right: firing probability as a function of photostimulation frequency. When all light pulses were considered, spike probability decreased with stimulation frequency (dark blue). However, the first 5 light pulses reliably evoke action potentials up to 100 Hz (light blue; spike probability = 0.78 at 100 Hz). **e**, Relative light-induced

firing rate change (log scale) versus latency of the maximal effect (peak/trough of PSTH). Three separate groups are apparent in mPFC: short-latency activated (VIP, green), inhibited (purple) and delayed activated (light brown). **f**, **g**, Top left: baseline firing rate versus spike width in ACx (**f**) and mPFC (**g**). Top right: cumulative fraction of firing rate. Bottom: cumulative fraction of spike width. Green, directly activated (VIP); purple, inhibited; light brown, delayed activated group; dark grey, unidentified neurons. Light grey depicts neurons for which inhibition could not reliably be assessed because of very low baseline firing rates (see Methods).
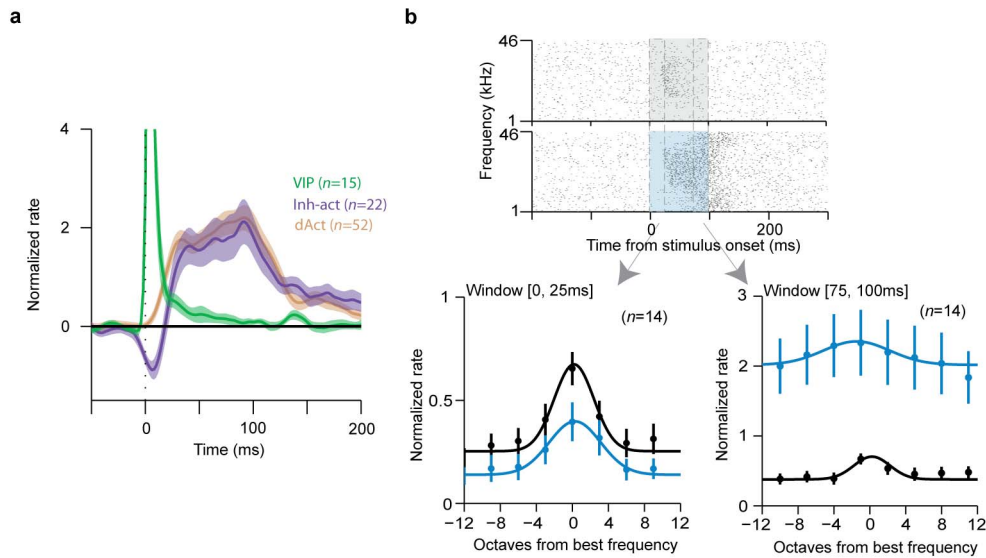
Extended Data Figure 3 | Inhibited and activated (Inh-act) population is a subgroup of the inhibited neurons. **a**, Average PSTH aligned to photostimulation (1-ms pulses) for Inh-act cells (a subgroup of the inhibited neurons, Figs 1 and 3). The colour code of Fig. 1g, h applies. Inh-act neurons (purple) show initial inhibition followed by delayed activation after 1-ms pulses. **b**, Top: example raster plot aligned to auditory stimuli of a tone-responsive Inh-act neuron. Shading indicates the stimulation windows (grey, tone only; blue, tone- and photostimulation). Dashed boxes indicate time windows for frequency tuning analysis (early, 0–25 ms; late, 75–100 ms). Bottom: frequency tuning curves of tone-responsive Inh-act neurons (population average, $n = 14$). Bottom left: tuning curve for the early time window (0–25 ms). Bottom right: tuning curve for the late time window (75–100 ms). Simultaneous photostimulation (100 ms) decreased the tone-evoked firing rates of Inh-act neurons in the early time window, whereas it increased the firing rates in the late time window. This pattern resembled the inhibition-activation sequence elicited by the 1 ms light pulses (**a**).
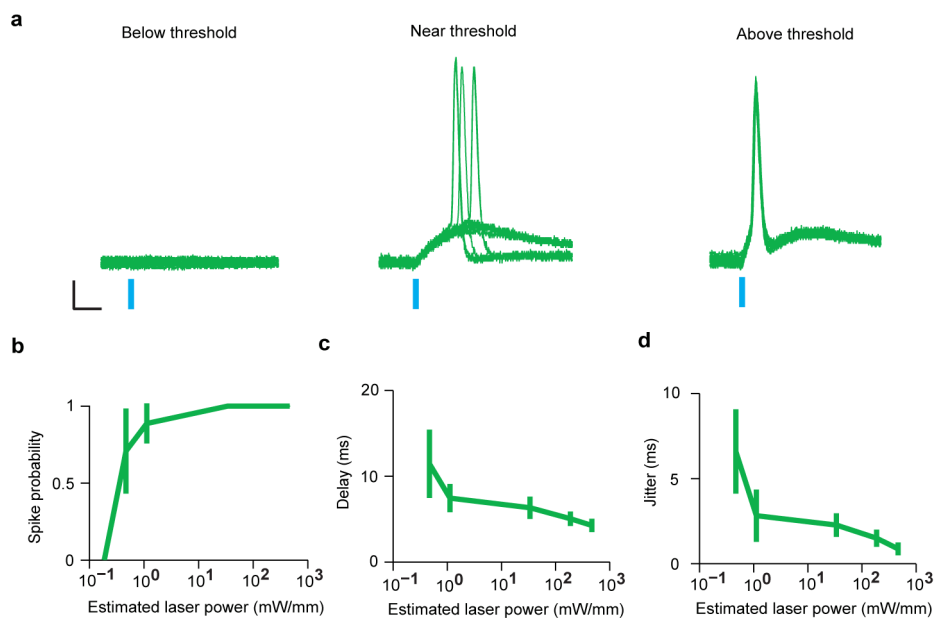
Extended Data Figure 4 | Light-intensity-dependent changes in spike probability, delay and jitter in VIP interneurons *in vitro*. a, Example traces of action potentials evoked by different light intensities. Blue bar, light stimulation. Scale bar, 10 ms, 10 mV. **b–d**, Quantification. Spike probability (**b**) increased whereas delay (**c**) and jitter (**d**) decreased with increasing light intensities. The highest two intensities were used in the *in vitro* experiments.
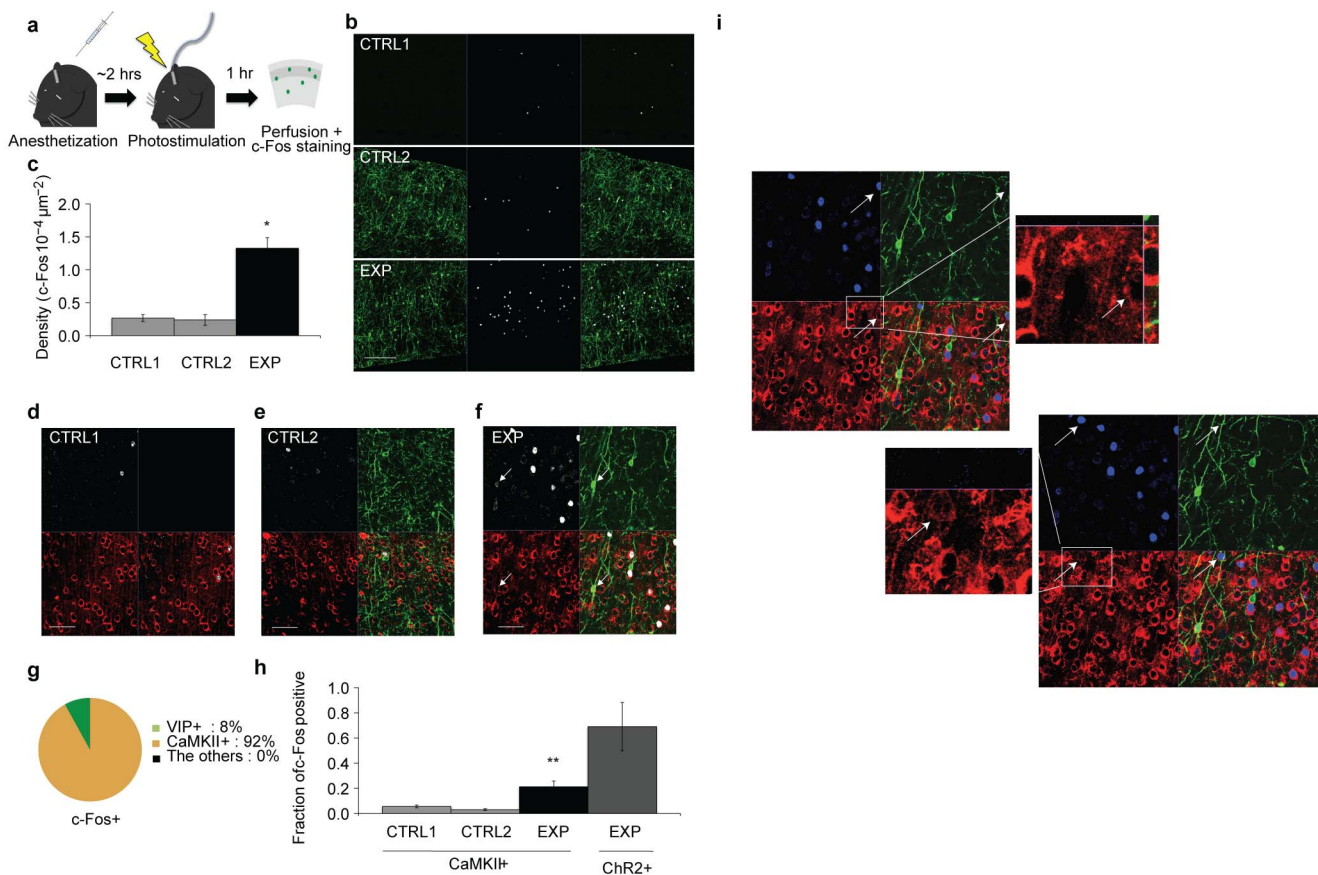
Extended Data Figure 5 | Photostimulation of VIP increases c-Fos in pyramidal neurons in mPFC. a, Schematic of c-Fos experiment. Animals were anaesthetized for 2 h to reduce the background c-Fos levels and photostimulation was applied. The expression level of c-Fos was captured 1 h after photostimulation. **b**, Representative images of different experimental conditions. CTRL1, no ChR2 expression with photostimulation; CTRL2, ChR2 expression without photostimulation; EXP, ChR2 expression with photostimulation. Left column: green, expression of ChR2.YFP; middle: white, c-Fos staining; right: merged images. Scale bar, 200 μm. **c**, Quantification of c-Fos levels. CTRL1, $n = 64$ c-Fos immunopositive neurons, $n = 4$ mice; CTRL2, $n = 58$ neurons, $n = 4$ mice; EXP, $n = 252$ neurons, $n = 4$ mice. **d–f**, Representative images from different experimental conditions. White, c-Fos staining; green, ChR2 expression; red, CaMKIIα staining. The arrow indicates a c-Fos-immunopositive VIP neuron. Scale bar, 50 μm. Note that some overlapping signals (c-Fos and CaMKIIα) are hard to appreciate in this image due to low resolution and uneven immunostaining. Additionally, owing to the different signal strength, CaMKIIα immunopositivity is hard to appreciate for neurons that are slightly above or below the focal plane, whereas the strong c-Fos immunoreactivity is still detectable. For this reason, additional high-power images were presented in **i**. **g**, Co-localization of markers with c-Fos. **h**, Fraction of c-Fos-positive cells. Among CaMKIIα-positive neurons, the proportion of c-Fos-immunopositive cells was significantly higher in the experimental group as compared with the controls. CTRL1, fraction = 11/208 (c-Fos/CaMKIIα), 4 mice; CTRL2, fraction = 3/107, 3 mice; EXP, fraction = 66/257, 4 mice. Approximately 64% (7 of 11) of the ChR2-expressing neurons were c-Fos-immunopositive. **i**, High-resolution images of the co-localization between c-Fos and CaMKIIα in **f**. Owing to low resolution and uneven staining of CaMKIIα, some c-Fos signals seemingly do not co-localize with CaMKIIα-positive neurons. However, in high-resolution images, the co-localization is clearer. Top: example of a weakly stained CaMKIIα-positive neuron (arrow). In the high-resolution image, CaMKIIα staining is apparent. Bottom: attributed to differences in immunofluorescence strength between c-Fos and CaMKIIα, neurons slightly out of focus may appear c-Fos-positive and CaMKIIα-negative. However, when the focal plane was adjusted, the co-localization became apparent. Blue, c-Fos; green, ChR2–YFP; red, CaMKIIα.
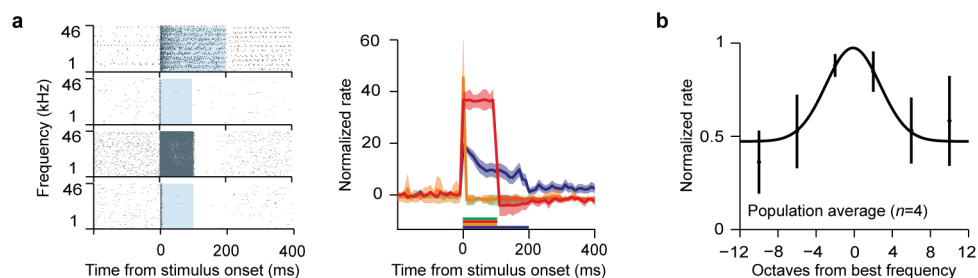
Extended Data Figure 6 | Responses of VIP neurons in ACx during auditory stimulation, alone or combined with photostimulation. a, Raster plot (left) and PSTH (right) aligned to the onset of combined tone and photostimulation. Although all VIP neurons were responsive to photostimulation, individual VIP neurons showed heterogeneous response profiles. One VIP neuron (top) showed accommodation during the 200-ms stimulation; two neurons (second and fourth) showed transient response; one neuron (third) fired persistently throughout the stimulation. Shaded boxes (left) or coloured lines (right) indicate the stimulation duration. **b**, Average frequency tuning curve of VIP neurons ($n = 4$). All 4 VIP neurons in the ACx for which tuning curves were recorded were responsive to pure tones; however, their tuning properties showed considerable heterogeneity.
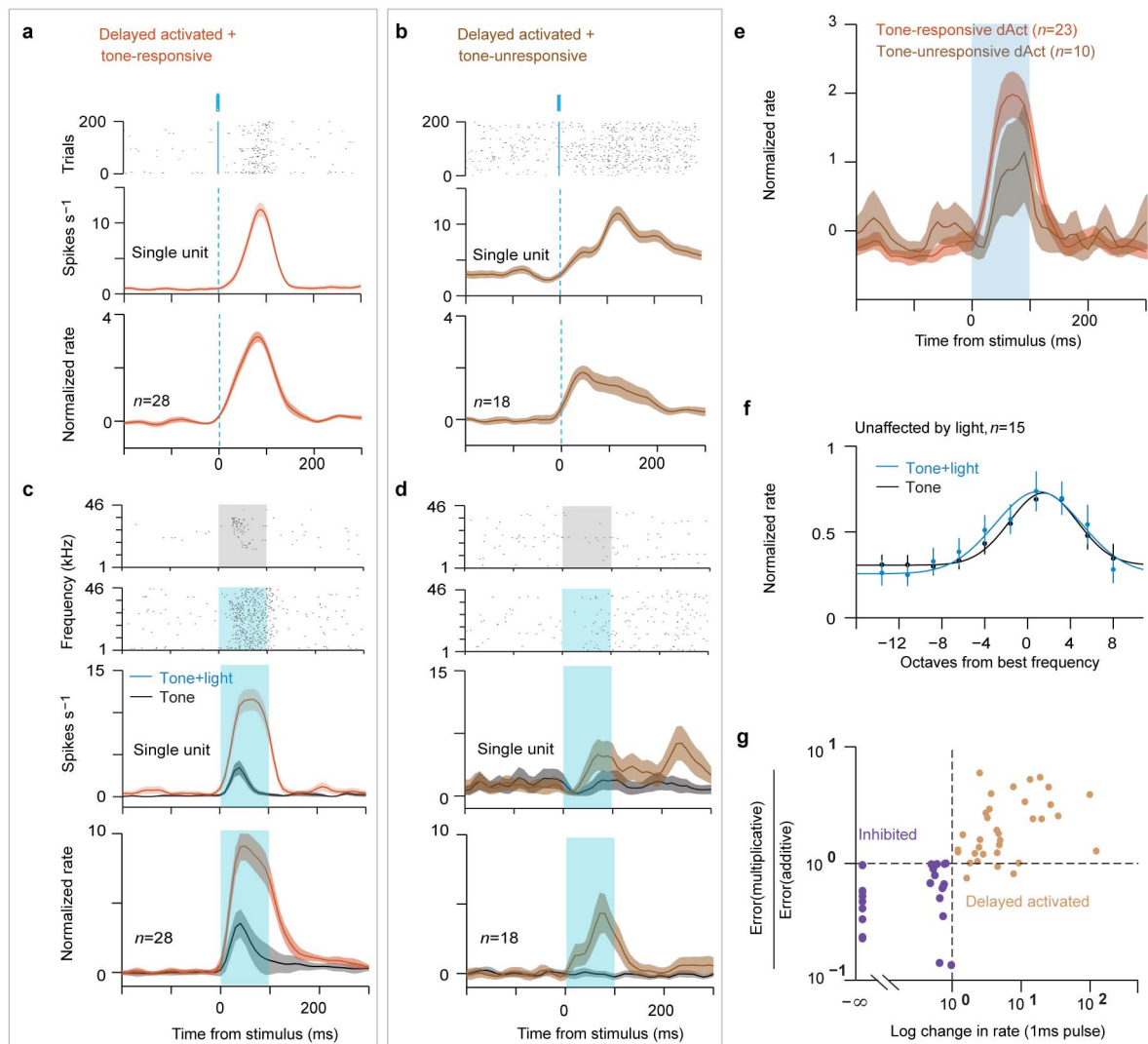
Extended Data Figure 7 | Auditory response profiles of different neuronal groups. a, b, Single-cell examples (top, raster plot; middle, PSTH) and population average (bottom, PSTH) of responses evoked by a brief 1-ms light pulse. **a**, Delayed activated and tone-responsive neurons. **b**, Delayed activated and tone-unresponsive neurons. **c, d**, Single-cell examples (top, raster plot; middle, PSTH) and population average (bottom, PSTH) of responses evoked by combined auditory and light stimulation (100 ms). **c**, Delayed activated and tone-responsive neurons. **d**, Delayed activated and tone-unresponsive neurons. Grey shaded box, tone stimulation; blue shaded box, tone plus light stimulation. **e**, PSTH of tone-responsive (red) and tone-unresponsive (brown) delayed activated neurons for 100-ms light pulses (without auditory stimulation). This experiment was performed in a subset of the frequency tuning experiments.

f, Frequency tuning curve of tone-responsive neurons not modulated by 1-ms light pulses. **g**, Fitting of one-parameter gain control models on tuning curve modulation of inhibited and delayed activated neurons. Tuning curves recorded during photostimulation were fitted with one-parameter models representing the scaled (multiplicative model) or shifted (additive model) versions of the baseline tuning curves (that is, without photostimulation). The ratio of the least squared errors of the two model fits is plotted as a function of relative firing rate change after 1-ms light pulses on a logarithmic scale, for inhibited (purple) and delayed activated (brown) neurons (minus infinity corresponds to complete abolishment of firing). An error ratio >1 corresponds to a better fit of the additive model, whereas <1 means better fit of the multiplicative gain model. See also Methods.
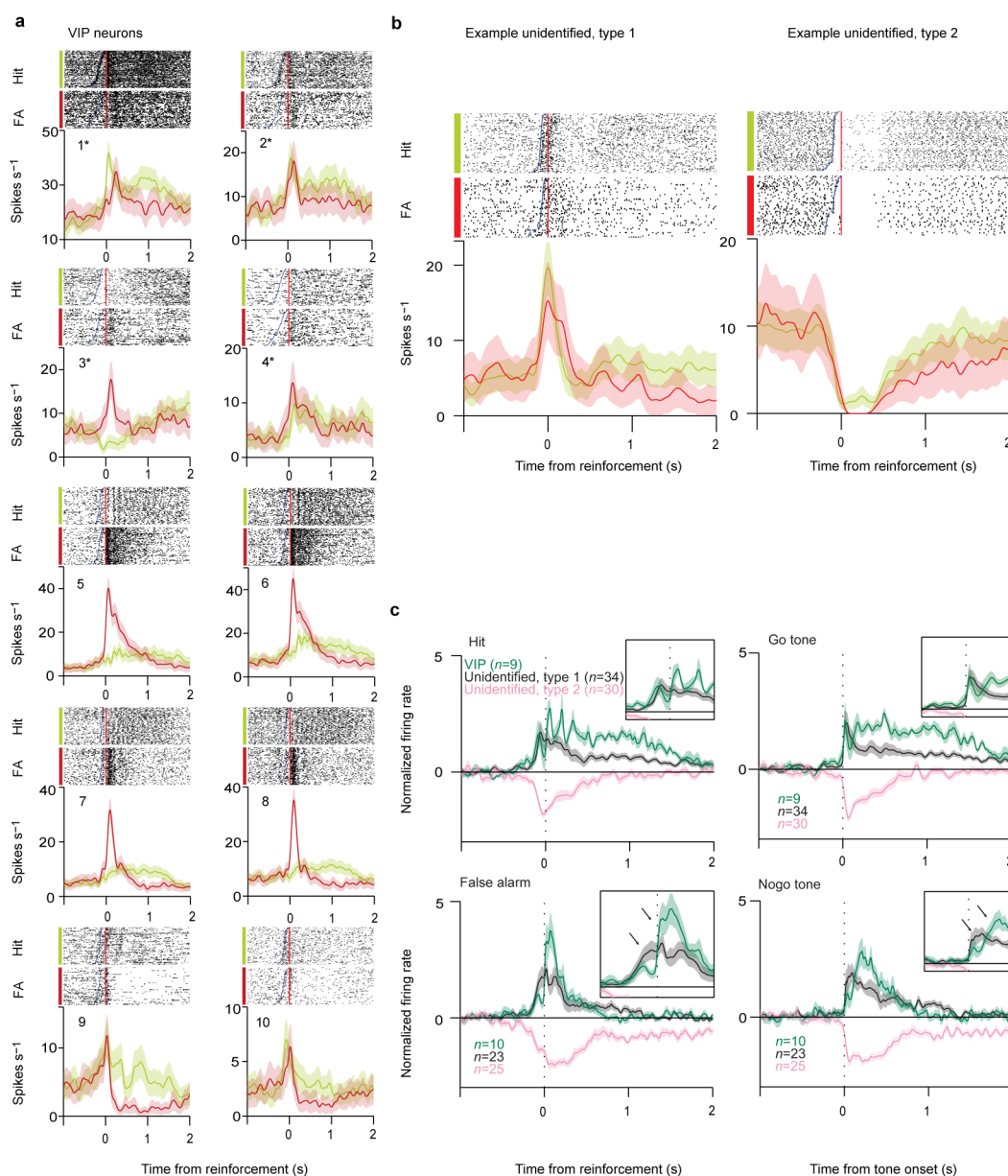
Extended Data Figure 8 | VIP neurons strongly respond to punishment in ACx. a, Raster plots (top) and PETHs (bottom) aligned to reinforcement (reward, green or punishment, red) for all VIP neurons recorded in the auditory go/no-go task. All VIP neurons were strongly recruited by punishment (foot shock: 1–4, marked by asterisks; air puff: 5–10), whereas water reward induced weaker and more sustained activation (9 of 10 cells showed significant firing rate increase for reward, see main text and Methods). **b**, Raster plots and PETHs of example unidentified neurons. Type 1 neurons (left) tended to be activated by tone onset. Type 2 neurons (right) tended to be inhibited by tone onset. **c**, Top: normalized average PETH of VIP (green) and concurrently recorded non-VIP neurons aligned to feedback (left) and tone onset (right) for hit (top) and false alarm trials (bottom). VIP neurons showed an abrupt increase of firing rate after punishment (bottom left). The oscillatory pattern of VIP activation around reward delivery is a consequence of rhythmic firing-rate modulations following the highly stereotypic pattern of licking in 4 of 10 VIP neurons (1 of 130 in non-VIP neurons). Grey, feedback-activated unidentified neurons (type 1). Pink, feedback-inhibited unidentified neurons (type 2). Insets, zoomed-in plots of PETHs. Arrows indicate the difference in activation pattern between VIP and unidentified type 1 neurons.
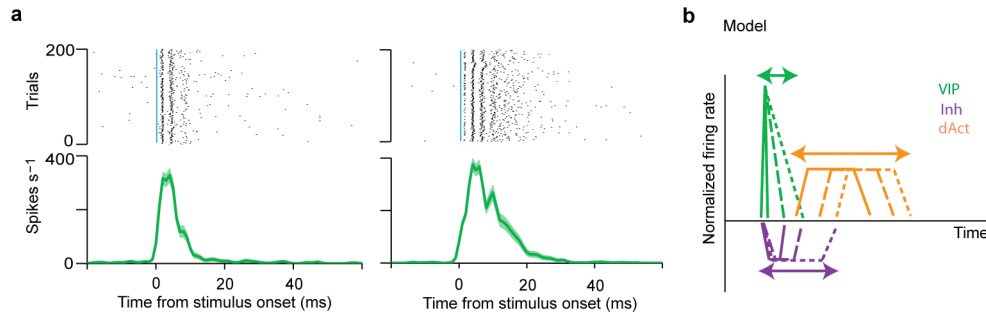
Extended Data Figure 9 | Accounting for the temporal difference between VIP-Cre::Ai32 (ACx) and VIP-Cre::AAV.ChR2 (mPFC). We observed a temporal difference in the firing pattern of VIP neurons between the ChR2 reporter line (Ai32) and the virus-injected (AAV.ChR2) mice. We speculated that this difference could stem from the mutation in ChR2. The mutation in ChR2(H134R) of Ai32 mice produces larger currents and slower kinetics than ChR2(ChETA; AAV.ChR2). As a consequence, VIP neurons in VIP-Cre::Ai32 can fire bursts in response to single 1-ms pulses and the activation can last more than 20 ms (**a**). This sustained activity of VIP neurons prolonged the temporal dynamics of downstream neurons. **a**, Examples of VIP neurons that burst to 1-ms photostimulation in the ACx of VIP-Cre::Ai32 mice. **b**, Model explaining the temporal difference between VIP-Cre::Ai32 and VIP-Cre::AAV-DIO-ChR2(ChETA). In VIP-Cre::Ai32 mice, almost all VIP neurons express ChR2 and exert stronger inhibition on the inhibited neuron group. Because the duration of VIP (green) activation varies, individual inhibited neurons (Inh, purple) receive different degrees of inhibition (strength and duration) from VIP neurons, therefore their firing rates recover to baseline at different time points. This variation propagates to the delayed activated group (dAct, orange), activation of which can start at different time instances.

Extended Data Table 1 | Contingency table showing delayed light activation and tone responsiveness are not independent

a. All neurons

	Tone-responsive	Tone-unresponsive
Delayed activated	28	18
Not delayed activated	69	227

Fisher's exact test : $P = 0.000001$

b. Putative pyramidal neurons

	Tone-responsive	Tone-unresponsive
Delayed activated	28	18
Not delayed activated (Spike width > 275µs)	39	174

Fisher's exact test : $P = 0.00000002$

LETTER

Differential L1 regulation in pluripotent stem cells of humans and apes

Maria C. N. Marchetto[1]*, Iñigo Narvaiza[1]*, Ahmet M. Denli[1], Christopher Benner[1], Thomas A. Lazzarini[1], Jason L. Nathanson[2], Apuã C. M. Paquola[1], Keval N. Desai[3], Roberto H. Herai[4], Matthew D. Weitzman[5], Gene W. Yeo[2], Alysson R. Muotri[4,6] & Fred H. Gage[1,6]

Identifying cellular and molecular differences between human and non-human primates (NHPs) is essential to the basic understanding of the evolution and diversity of our own species. Until now, preserved tissues have been the main source for most comparative studies between humans, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*)[1,2]. However, these tissue samples do not fairly represent the distinctive traits of live cell behaviour and are not amenable to genetic manipulation. We propose that induced pluripotent stem (iPS) cells could be a unique biological resource to determine relevant phenotypical differences between human and NHPs, and that those differences could have potential adaptation and speciation value. Here we describe the generation and initial characterization of iPS cells from chimpanzees and bonobos as new tools to explore factors that may have contributed to great ape evolution. Comparative gene expression analysis of human and NHP iPS cells revealed differences in the regulation of long interspersed element-1 (L1, also known as LINE-1) transposons. A force of change in mammalian evolution, L1 elements are retrotransposons that have remained active during primate evolution[3–5]. Decreased levels of L1-restricting factors APOBEC3B (also known as A3B)[6] and PIWIL2 (ref. 7) in NHP iPS cells correlated with increased L1 mobility and endogenous L1 messenger RNA levels. Moreover, results from the manipulation of A3B and PIWIL2 levels in iPS cells supported a causal inverse relationship between levels of these proteins and L1 retrotransposition. Finally, we found increased copy numbers of species-specific L1 elements in the genome of chimpanzees compared to humans, supporting the idea that increased L1 mobility in NHPs is not limited to iPS cells in culture and may have also occurred in the germ line or embryonic cells developmentally upstream to germline specification during primate evolution. We propose that differences in L1 mobility may have differentially shaped the genomes of humans and NHPs and could have continuing adaptive significance.

Humans, chimpanzees and bonobos are genetically very similar, sharing nearly 98% of their alignable genomic sequence[1–3]. However, cellular and molecular phenotypes, especially at identical stages of development, are difficult to establish, mainly owing to limited access to embryonic material from humans and NHPs[8]. We reprogrammed fibroblasts from two bonobos and two chimpanzees into iPS cells as previously described[9,10] (Extended Data Fig. 1a). After culture in human embryonic stem (ES) cell-supporting conditions, NHP iPS cell colonies could be distinguished by the high nucleus-to-cytoplasm ratio morphology. iPS cell clones from both species continuously expressed pluripotency markers, retained an undifferentiated morphology in culture, and maintained a normal karyotype (Fig. 1a). After
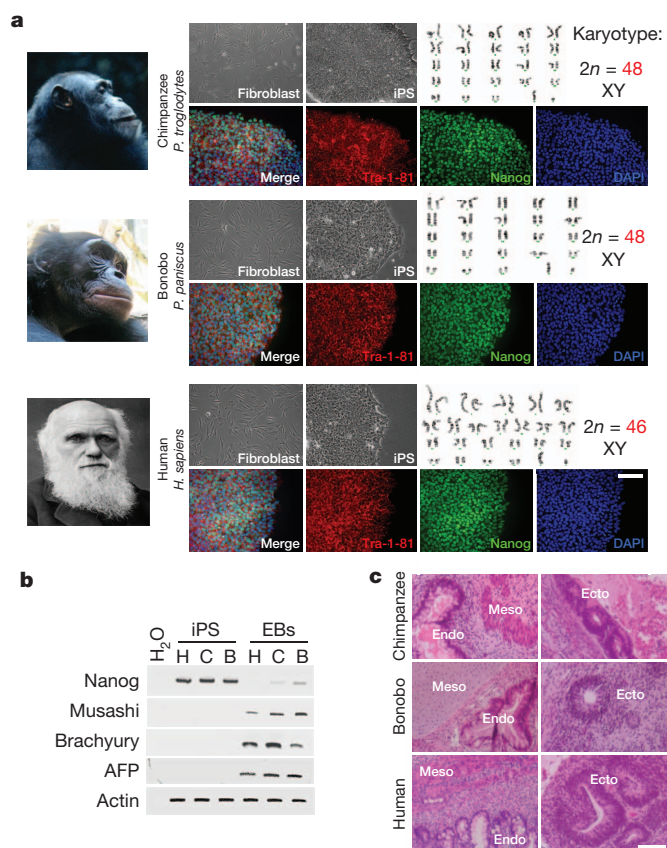
Figure 1 | **Characterization of iPS cells derived from the three primate species. a,** Morphology of fibroblasts and iPS cells. No karyotypic abnormalities were observed in iPS cells clones. Immunofluorescence for the pluripotency markers Tra-1-81 and Nanog in iPS cells is shown. DAPI, 4′,6-diamidino-2-phenylindole. **b,** Reverse transcription PCR (RT–PCR) for undifferentiation (Nanog) and for the three germ cell layers (musashi, brachyury and α-fetoprotein (AFP)) markers in human (H), chimpanzee (C) and bonobo (B) iPS cells, and in differentiated embryoid bodies (EBs). **c,** Haematoxylin and eosin staining of teratoma sections showing differentiation into three germ layers: goblet cells in gastrointestinal tract (endo), neuroretinal epithelium (ecto), and muscle and cartilage/bone (meso). Scale bars, 100 μm (**a**) and 150 μm (**c**). Portrait of Charles Darwin is reproduced with permission of The Huntington Library.

[1]Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [2]University of California San Diego, Department of Cellular and Molecular Medicine, Stem Cell Program, Institute for Genomic Medicine, Sanford Consortium for Regenerative Medicine, 2880 Torrey Pines Scenic Drive, La Jolla, California 92037, USA. [3]University of California San Diego, Division of Biological Sciences, 9500 Gilman Drive, La Jolla, California 92093, USA. [4]University of California San Diego, School of Medicine, Department of Pediatrics/Rady Children's Hospital San Diego, Department of Cellular & Molecular Medicine, Stem Cell Program, Sanford Consortium, MC 0695, 2880 Torrey Pines Scenic Drive, La Jolla, California 92093, USA. [5]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine and Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104-4318, USA. [6]Center for Academic Research and Training in Anthropogeny (CARTA), 9500 Gilman Drive, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

embryoid-body-mediated differentiation *in vitro*, clones contained tissue derivatives from the three embryonic germ layers and down-regulated expression of pluripotency markers (Fig. 1b). iPS-cell-selected clones were also able to differentiate into the three embryonic germ layers *in vivo*, as shown by analysis of teratomas in nude mice (Fig. 1c). Together, these data demonstrate that NHP iPS cell clones re-established pluripotency at the molecular and cellular levels.

To gain insight into differences in gene expression between human and NHP iPS cells, we performed high-throughput RNA sequencing (RNA-seq) analyses on four human, two chimpanzee and two bonobo iPS cell lines (Extended Data Fig. 1b). The expression profiles of iPS cells from the three species clustered together with human ES cells (HUES6 and H1), and were distinguishable from ES-cell-derived neural progenitor cells (Fig. 2a); chimpanzee and bonobo iPS cells clustered closer to each other than to human iPS cells (Fig. 2a). We then per-formed pairwise comparisons of protein-coding gene expression levels (Fig. 2b). Venn diagrams in Fig. 2b represent expressed genes with non-significant differences between species (purple), and upregulated genes with estimated false discovery rates (FDR) of less than 5% and a fold change greater than twofold (pale orange and blue). Comparison between humans and NHPs (Fig. 2b, bottom right) revealed 1,376 genes with increased expression in human iPS cells, and 1,042 common

genes with increased expression in NHP iPS cells, whereas no signifi-cant differences were observed in 11,585 protein-coding genes. Next, we focused on genes differentially expressed between human and NHP iPS cells (Fig. 2c–e and Extended Data Fig. 1c, d), and found, among the top 50 genes with increased expression in human compared to NHP iPS cells, two genes involved in the restriction of L1 retrotransposition, namely *A3B* and *PIWIL2* (Fig. 2d).

Active, full-length L1 elements have the ability to move from one location in the genome to another by a copy–paste mechanism known as retrotransposition[11]. Active L1 elements have been detected in both germline and somatic tissues, and can affect genome integrity[12,13]. As uncontrolled retrotransposition activity can be deleterious to the host[14], organisms have evolved mechanisms to control L1 mobility[11]. A3B is a member of the APOBEC3 family of cytidine deaminases that can inhibit L1 mobility in different cell types, including human ES and iPS cells, via a still unclear mechanism[6,15,16]. PIWIL2 is an effector of the Piwi-interacting RNA (piRNA) pathway involved in L1 silencing mainly in the germ line[7].

To confirm differences in *A3B* and *PIWIL2* in human versus NHP iPS cells, we first cloned their complementary DNAs from the three species, and found a high degree of conservation between humans and NHPs (Extended Data Fig. 2). Quantification of *A3B* mRNA levels by
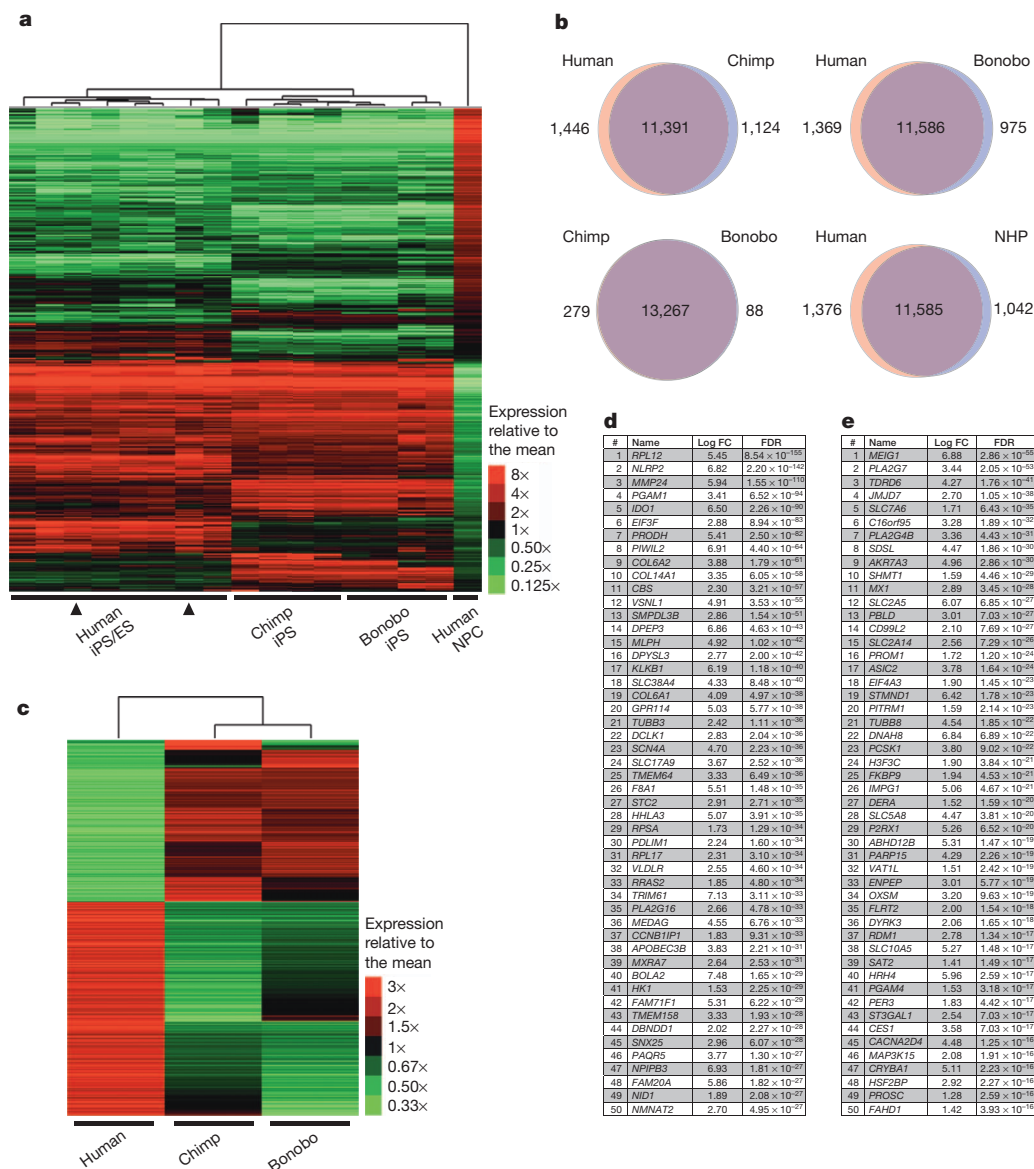
Figure 2 | RNA-seq profiling of human and NHP iPS cells. a, High-throughput sequencing of 14 RNA samples corresponding to four human, two chimp and two bonobo iPS cell lines. Expression profiles of human ES cells (H1 and HUES6, arrowheads) and ES-cell-derived neural progenitor cells (NPCs) are shown. Heat-map representation of mapped reads corresponds to protein-coding genes. **b**, Venn diagrams showing pairwise comparison of protein-coding genes. Pale orange and blue denote significantly upregulated genes (FDR < 0.05 and fold change > 2); purple denotes expressed genes with no significant differences in mRNA levels between compared species. **c**, Heat-map representation of differentially expressed protein-coding genes with FDR < 0.05 and fold change > 2 between human and NHP iPS cells. **d, e**, List of the top 50 differentially expressed genes in human compared to NHP iPS cells, with increased expression in human (**d**) and NHP (**e**) iPS cells. *PIWIL2* and *APOBEC3B* were expressed at significantly higher levels in human than in NHP iPS cells (positions 8 and 38 in **d**, respectively). Rank, gene name, logarithmic base 2 fold change (log FC), and FDR are shown.

d

#	Name	Log FC	FDR
1	RPL12	5.45	8.54×10^{-155}
2	NLRP2	6.82	2.20×10^{-142}
3	MMP24	5.94	1.55×10^{-110}
4	PGAM1	3.41	6.52×10^{-94}
5	IDO1	6.50	2.26×10^{-90}
6	EIF3F	2.88	8.94×10^{-83}
7	PRODH	5.41	2.50×10^{-82}
8	PIWIL2	6.91	4.40×10^{-64}
9	COL6A2	3.88	1.79×10^{-61}
10	COL14A1	5.35	6.05×10^{-58}
11	CBS	2.30	3.21×10^{-57}
12	VSNL1	4.91	3.53×10^{-55}
13	SMPDL3B	2.86	1.54×10^{-51}
14	DPEP3	6.86	4.63×10^{-45}
15	MLPH	4.92	1.02×10^{-42}
16	DPYSL3	2.77	2.00×10^{-42}
17	KLKB1	6.19	1.18×10^{-40}
18	SLC38A4	4.33	8.48×10^{-40}
19	COL6A1	4.09	4.97×10^{-38}
20	GPR114	5.03	5.77×10^{-38}
21	TUBB3	2.42	1.11×10^{-36}
22	DCLK1	2.83	2.04×10^{-36}
23	SCN4A	4.70	2.23×10^{-36}
24	SLC17A9	3.67	2.52×10^{-36}
25	TMEM64	3.33	6.49×10^{-36}
26	F8A1	5.51	1.48×10^{-35}
27	STC2	2.91	2.71×10^{-35}
28	HHLA3	5.07	3.91×10^{-35}
29	RPSA	1.73	1.29×10^{-34}
30	PDLIM1	2.24	1.60×10^{-34}
31	RPL17	2.31	3.10×10^{-34}
32	VLDLR	2.55	4.60×10^{-34}
33	RRAS2	1.85	4.80×10^{-35}
34	TRIM61	7.13	3.11×10^{-33}
35	PLA2G16	2.66	4.78×10^{-33}
36	MEDAG	4.55	6.76×10^{-33}
37	CCNB1IP1	1.83	9.31×10^{-33}
38	APOBEC3B	3.83	2.21×10^{-31}
39	MXRA7	2.64	2.53×10^{-31}
40	BOLA2	7.48	1.65×10^{-30}
41	HK1	1.53	2.25×10^{-29}
42	FAM71F1	5.31	6.22×10^{-29}
43	TMEM158	3.33	1.93×10^{-28}
44	DBNDD1	2.02	2.27×10^{-28}
45	SNX25	2.96	6.07×10^{-28}
46	PAQR5	3.77	1.30×10^{-27}
47	NPIPB3	6.93	1.81×10^{-27}
48	FAM20A	5.86	1.82×10^{-27}
49	NID1	1.89	2.08×10^{-27}
50	NMNAT2	2.70	4.95×10^{-27}

e

#	Name	Log FC	FDR
1	MEIG1	6.88	2.86×10^{-55}
2	PLA2G7	3.44	2.05×10^{-53}
3	TDRD6	4.27	1.76×10^{-41}
4	JMJD7	2.70	1.05×10^{-38}
5	SLC7A6	1.71	6.43×10^{-35}
6	C16orf95	3.28	1.89×10^{-32}
7	PLA2G4B	3.36	4.43×10^{-31}
8	SDSL	4.47	1.86×10^{-30}
9	AKR7A3	4.96	2.86×10^{-30}
10	SHMT1	1.59	4.46×10^{-29}
11	MX1	2.89	3.45×10^{-29}
12	SLC2A5	6.07	6.85×10^{-27}
13	PBLD	3.01	7.03×10^{-27}
14	CD96L2	2.10	7.69×10^{-27}
15	SLC2A14	2.56	7.29×10^{-26}
16	PROM1	1.72	1.20×10^{-24}
17	ASIC2	3.78	1.64×10^{-24}
18	EIF4A3	1.90	1.45×10^{-23}
19	STMND1	6.42	1.78×10^{-23}
20	PITRM1	1.59	2.14×10^{-21}
21	TUBB8	4.54	1.85×10^{-22}
22	DNAH8	6.84	6.89×10^{-22}
23	PCSK1	3.80	9.02×10^{-22}
24	H3F3C	1.90	3.84×10^{-21}
25	FKBP9	1.94	4.53×10^{-21}
26	IMPG1	5.06	4.67×10^{-21}
27	DERA	1.52	1.59×10^{-20}
28	SLC5A8	4.47	3.81×10^{-20}
29	P2RX1	5.26	6.52×10^{-20}
30	ABHD12B	5.31	1.47×10^{-19}
31	PARP15	4.29	2.26×10^{-19}
32	VAT1L	1.51	2.42×10^{-19}
33	ENPEP	3.01	5.77×10^{-19}
34	OXSM	3.20	9.63×10^{-19}
35	FLRT2	2.00	1.54×10^{-18}
36	DYRK3	2.06	1.65×10^{-18}
37	RDM1	2.78	1.34×10^{-17}
38	SLC10A5	5.27	1.48×10^{-17}
39	SAT2	1.41	1.49×10^{-17}
40	HRH4	5.96	2.59×10^{-17}
41	PGAM4	1.53	3.18×10^{-17}
42	PER3	1.83	4.42×10^{-17}
43	ST3GAL1	2.54	7.03×10^{-17}
44	CES1	3.58	7.03×10^{-17}
45	CACNA2D4	4.48	1.25×10^{-16}
46	MAP3K15	2.08	1.91×10^{-16}
47	CRYBA1	5.11	2.23×10^{-16}
48	HSF2BP	2.92	2.27×10^{-16}
49	PROSC	1.28	2.59×10^{-16}
50	FAHD1	1.42	3.93×10^{-16}

Figure 3 | Reduced levels of *A3B* and *PIWIL2* and increased L1 mobility in NHP iPS cells. a, **b**, qPCR analysis of *A3B* (**a**) and *PIWIL2* (**b**) expression in human and NHP iPS cells (Extended Data Figs 2 and 9). **c**, Immunoblot for A3B and PIWIL2. **d**, Effect of A3B and PIWIL2 on L1-expressing firefly luciferase (L1-Luc) retrotransposition in 293T cells. 293T cells were co-transfected with L1-luciferase plasmid (pYX017)[19] plus control (ctrl), PIWIL2-, A3B- or A3G-expressing plasmid. L1-luciferase mobility was calculated as firefly luciferase units relative to *Renilla* luciferase units. L1 activity is shown relative to control. **e**, Comparable levels of L1–eGFP[18] mobility in human ES (hES) and iPS (hiPS) cells. L1–eGFP mobility is shown as a percentage of eGFP-positive cells by fluorescence-activated cell sorting (FACS) relative to human ES cells. **f**, L1–eGFP retrotransposition in human, chimp and bonobo iPS cells. L1 mobility was calculated as a percentage of eGFP-positive cells and shown as relative L1 mobility to human iPS cell line 1 (iPS1). **g**, Representative images of human, chimpanzee and bonobo iPS cells transfected with L1–eGFP. Scale bar, 50 μm. **h**, Retrotransposition quantification of species-specific L1 elements. The mobility of human and chimp reporter L1–eGFP elements (human-L1 and chimp-L1, respectively) was quantified in transfected human, chimp and bonobo iPS cells. Retrotranposition activity is shown relative to human-L1 activity in human iPS cells. Error bars denote s.e.m. *$P < 0.01$ between indicated groups using *t*-test ($n = 3$ (**a**, **b**, **e** and **f**) and 4 (**d** and **h**) biological replicates).

quantitative PCR (qPCR) confirmed significantly higher levels (~30-fold) of *A3B* in both human iPS cell lines compared to NHP iPS cells (Fig. 3a). Levels of *PIWIL2* mRNA were 16-fold higher in human than in NHP iPS cell lines (Fig. 3b). PIWIL2-mediated control of transposons is most active in the germ line, and we observed that levels of *PIWIL2* mRNA are 20–40-fold lower in human iPS cells than in the testis (Extended Data Fig. 3a). The increased expression observed in human iPS cells seems to be specifically restricted to *A3B* and *PIWIL2* compared to other members of these protein families (Extended Data Fig. 3b, c). Differences in *A3B* and *PIWIL2* mRNA levels reflected higher A3B and PIWIL2 protein levels in human versus NHP iPS cells (Fig. 3c).

Ectopic expression of A3B has been shown to inhibit the mobility of human L1 reporter elements[6,17–19] (Extended Data Fig. 4a). In 293T cells, ectopic expression of human A3B significantly reduced L1-expressing firefly luciferase[18,19] mobility by fivefold compared to control plasmid or a plasmid expressing A3G, another APOBEC3 protein that lacks anti-L1 activity[17] (Fig. 3d). We also found a significant decrease in L1-luciferase retrotransposition in cells overexpressing PIWIL2 compared to control transfected cells (Fig. 3d). We then confirmed that human L1 can retrotranspose in human ES and iPS cells under our

culture conditions, as previously shown[15,20,21] (Fig. 3e). Because we found reduced levels of L1 restriction factors A3B and PIWIL2 in NHPs, we compared L1 activity in human versus NHP iPS cells using human L1 tagged with the enhanced green fluorescent protein (eGFP) reporter element[18,19]. L1 retrotransposition was significantly higher in NHP compared to human iPS cell lines, with 10- and 8-fold increases in eGFP-positive cells in chimpanzee and bonobo iPS cells, respectively (Fig. 3f, g). To test whether the differential L1 regulation in iPS cells is specific to human L1, we measured the activity of an NHP L1 element in iPS cells. We generated a retrotransposition-competent chimpanzee L1–eGFP reporter element (chimp-L1) (Extended Data Fig. 5), and observed that chimp-L1 was significantly more active in NHP than in human iPS cells (Fig. 3h), suggesting that the decreased L1 activity in human iPS cells is not specific to the human L1 element, and that human iPS cells are more efficient in repressing L1 retrotransposition than NHP iPS cells.

To analyse the contribution of endogenous A3B to the differential L1 activity observed among primate cells, we generated human ES and iPS cells with decreased levels of A3B (Fig. 4a and Extended Data Fig. 4b, c). Stable short hairpin RNA (shRNA)-mediated knock-down of A3B (shA3B) resulted in a significant increase in L1–eGFP activity compared to scramble (shScr) control cells in human iPS cells (Fig. 4a, b). Knockdown was specific to A3B, not affecting other APOBEC3 proteins (Extended Data Fig. 4d–f). As expected, L1 mobility was significantly decreased in both chimpanzee and bonobo iPS cell lines after A3B overexpression (Fig. 4c). Similarly, overexpression of chimpanzee or bonobo PIWIL2 in NHP iPS cells reduced L1 mobility to levels detected in human iPS cells (Fig. 4d). We did not observe differences in the levels of L1–eGFP mRNA expressed from trans-fected L1 plasmid or in L1 promoter activity between human and NHP iPS cells (Extended Data Fig. 4g–i). Together, these results suggest that differences in A3B and PIWIL2 expression levels contribute to higher L1 retrotransposition in NHP than in human iPS cells.

PIWIL2 repression of transposons is mediated through piRNAs[22]. Thus, we analysed the presence of PIWIL2-bound piRNAs in doxycycline-inducible human iPS cells expressing Flag-tagged PIWIL2 by immunoprecipitation and subsequent 5′ end labelling. Analysis of PIWIL2-associated small RNAs revealed the presence of ~26–30-nucleotide RNAs only in cells expressing Flag–PIWIL2 but not in control cells or in pulldowns with control antibody (Fig. 4e and Extended Data 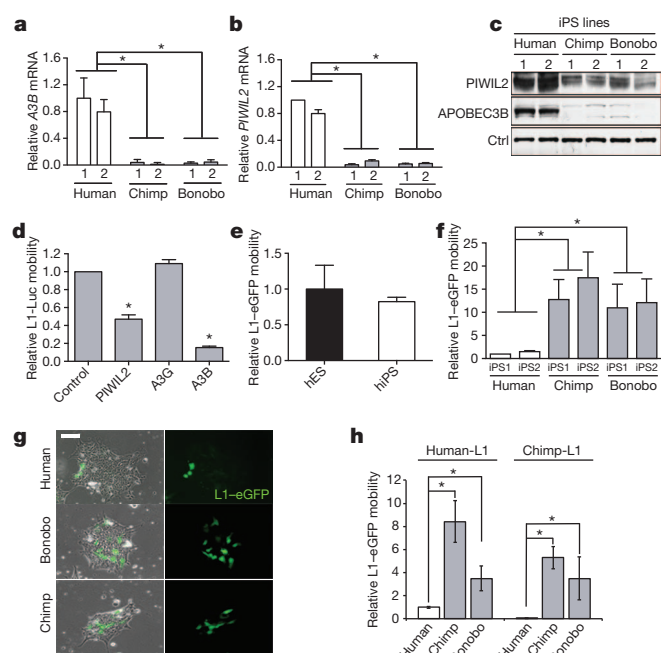Fig. 6a, b). Next, to probe for the presence of L1-targeting-piRNAs, we characterized the small RNA populations in human iPS cells by small RNA-seq analysis (Extended Data Fig. 6c, d and Supplementary Tables 1 and 2). We detected 272 and 229 annotated piRNAs in human iPS cell lines 1 and 2, respectively (Extended Data Fig. 6d–f and Supplementary Table 2). In addition, we observed a number of 26–33-nucleotide small RNAs mapping to the consensus human-specific L1 element (L1 *Homo sapiens*; L1Hs) sequence (Fig. 4 and Extended Data Fig. 7a, b), including 12 and 10 of the 37 annotated piRNAs mapping to L1Hs in piRNAbank (http://pirnabank.ibab.ac.in/) in hiPS cell lines 1 and 2, respectively (Extended Data Figs 6e, g and 7a–c). Together, these results demonstrate the presence of piRNAs complementary to L1Hs in human iPS cells.

We then asked whether different levels of L1 reporter mobility between human and NHP iPS cells reflect differences in endogenous L1 activity. First, we analysed endogenous L1 RNA levels by qPCR, and found higher levels of endogenous L1 mRNA in chimpanzee and bonobo than in human iPS cells (Fig. 4g and Extended Data Fig. 8a–c). Next, we examined the number of L1 elements in human and chimpanzee genomes to assess differences in recent L1 mobility. We did not observe major differences in the number of L1 elements for older families (L1PA4, L1AP3 and L1PA2; approximately 18, 12.5 and 7.6 million years old, respectively)[23,24] (Fig. 4h). However, we did observe significantly higher numbers of chimpanzee-specific L1 elements (L1 *Pan troglodytes*; L1Pt) compared to L1Hs elements[25,26] (Fig. 4h). Differences in the expression of *A3B* and *PIWIL2* suggest that L1 mobility may
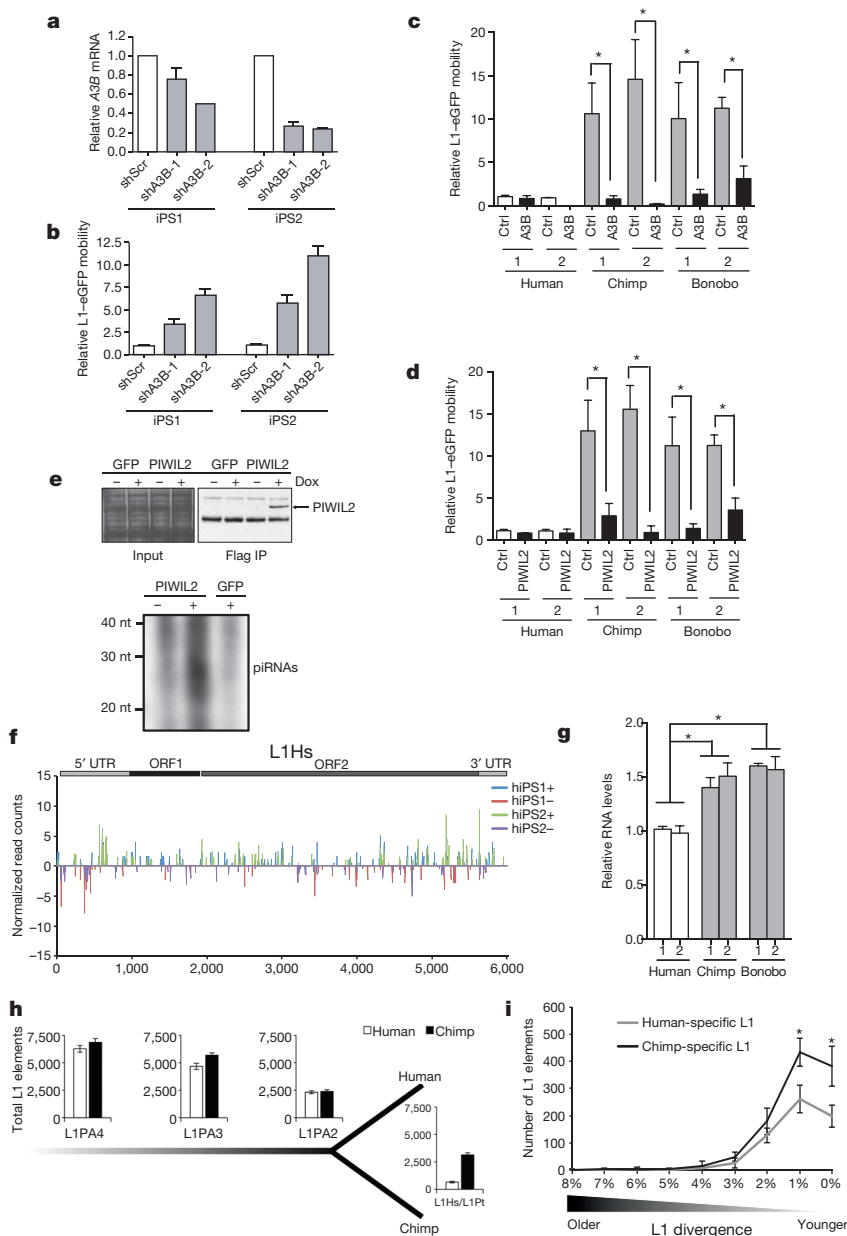
Figure 4 | Species-specific L1 elements are more abundant in chimpanzee genomes than in human genomes, correlating with decreased levels of A3B and PIWIL2. a, Stable shRNA-mediated knockdown of A3B (shA3B-1, shA3B-2) or control (shScr) in human iPS cells. *A3B* expression was normalized to *GAPDH* and shown relative to shScr. **b**, L1–eGFP mobility in shA3B iPS cells. eGFP-positive cells were quantified by FACS analysis and shown relative to shScr control. **c, d**, Overexpression of A3B (**c**) and PIWIL2 (**d**) decreases L1–eGFP retrotransposition in NHP iPS cells. Cells were electroporated with L1–eGFP plus control, A3B- or PIWIL2-expressing plasmids. L1–eGFP mobility is shown relative to human iPS cell-1 control. **e**, Immunoprecipitation (IP) of piRNAs associated with PIWIL2 in human iPS cells. Top, immunoprecipitation of PIWIL2 ribonucleoproteins (RNPs) from Tet-inducible GFP and Flag-tagged PIWIL2 human iPS cells after addition of doxycycline (dox). Bottom, $[\gamma\text{-}^{32}P]ATP$ 5′-end labelling of RNA associated with Flag–PIWIL2 RNPs. Size markers are indicated (nt, nucleotides). **f**, Mapping of 26–33-nucleotide RNA reads (containing uracil at the 5′ end and/or adenine at position +10) detected by small RNA-seq from human iPS cell lines 1 and 2 to consensus L1Hs (Repbase). Positive and negative values indicate sense and antisense piRNAs, respectively. Schematic representation of L1 is shown on top. Read counts were normalized to 10^7 reads per experiment. **g**, qRT–PCR analysis of endogenous L1 RNA in human and NHP iPS cells. Values represent average of relative levels for L1 RNA (5′ untranslated region (UTR), open reading frame (ORF) 1 and 2), normalized to *ACTB* mRNA. L1 levels are shown relative to iPS cell line 1. **h**, Comparative quantitative analysis of L1 elements in human and chimpanzee genomes for L1 families L1PA4, L1PA3, L1PA2, L1Pt and L1Hs. **i**, Number of species-specific L1 insertions (L1PA2, L1Hs and L1Pt) relative to their divergence. L1 elements plotted as a histogram relative to their divergence (number of mutations relative to the canonical element). Error bars denote s.d. (**h**) and s.e.m. (**a–d** and **g**). *$P < 0.001$ (**h**; between human and chimpanzee; Mann–Whitney U test) and *$P < 0.01$ (**c, d** and **g**; between indicated groups; t-test). $n = 3$ (**a–d**) and 4 (**g**) biological replicates.

have been altered at a relatively recent evolutionary divergence. Therefore, using divergence as a measurement of L1 age, we estimated the number of species-specific L1 loci, and found that the number of chimpanzee-specific loci was significantly higher than the number of human-specific loci (Fig. 4i and Extended Data Fig. 8d–g). This increased number of species-specific L1 loci in chimpanzee suggests that endogenous L1 has been more active in NHP genomes, correlating with the decreased levels of A3B and PIWIL2.

Here we show that iPS cells from both chimpanzees and bonobos have increased L1 mobility. Different rates of L1 activity could lead to considerable changes in genomic structure and function, and could potentially affect adaptation. The human population has gone through one or more bottlenecks throughout evolution that might have contributed to decreased genetic diversity[27]. Chimpanzees and bonobos, in contrast, have increased levels of genetic diversity when compared to humans[28]. This idea is also supported by data showing that there is substantially more genetic difference among individuals within chimpanzee troops in West Africa than among all living humans[29]. Although it remains unclear what the main generators of the phenotypic differences between us and our closest living relatives are (despite the

extreme genetic similarity), we propose that L1 mobility could be involved in differentially shaping the genomes of humans and NHPs, providing an extra layer of variability to the latter. In fact, recent studies have suggested that ongoing L1 retrotransposition may contribute to inter-individual genetic variation[30]. In this work we also present a new perspective on the use of iPS cell technology as a powerful tool for the study of early stages of development and possible validation of evolutionary genomic and transcriptomic modifications that identify humans as outliers among primates. The iPS cells from great apes that we describe here can also be used for comparative studies of any derivative pluripotent or terminally differentiated cell types, limited solely by the availability of differentiation protocols.

METHODS SUMMARY

Reprogramming of fibroblast was performed by transduction with retroviral vectors expressing *OCT4* (also known as *POU5F1*), *MYC*, *KLF4* and *SOX2* human cDNAs. For RNA-seq, libraries from polyA⁺ RNA and small RNA were generated using the Illumina TruSeq RNA and Small RNA TruSeq Sample Prep kits, respectively, and analysed on an Illumina HiSeq 2000 sequencer. L1 reporter assays were performed as previously described[18,19]. Quantification of L1-derived genomic

sequences was based on Repbase defined elements annotated by RepeatMasker (http://www.repeatmasker.org). L1 genomic positions for human (hg19, GRCh37) and chimpanzee (panTro3, CGSC 2.1.3) genomes were downloaded from the UCSC Genome Browser annotation database. To identify reference L1 elements that were inserted into the genome after the last common ancestor for human and chimpanzee, L1 elements were mapped between homologous regions of each genome using the UCSC LiftOver tool.

1. The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437,** 69–87 (2005).
2. Prüfer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486,** 527–531 (2012).
3. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
4. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
5. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303,** 1626–1632 (2004).
6. Bogerd, H. P. *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl Acad. Sci. USA* **103,** 8780–8785 (2006).
7. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316,** 744–747 (2007).
8. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* **15,** 1746–1758 (2005).
9. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
10. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
11. Burns, K. H. & Boeke, J. D. Human transposon tectonics. *Cell* **149,** 740–752 (2012).
12. Kazazian, H. H. Jr. Mobile elements and disease. *Curr. Opin. Genet. Dev.* **8,** 343–350 (1998).
13. Muotri, A. R., Marchetto, M. C., Coufal, N. G. & Gage, F. H. The necessary junk: new functions for transposable elements. *Hum. Mol. Genet.* **16,** R159–R167 (2007).
14. Beck, C. R., Garcia-Perez, J. L., Badge, R. M. & Moran, J. V. LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* **12,** 187–215 (2011).
15. Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V. & Greene, W. C. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J. Biol. Chem.* **286,** 36427–36437 (2011).
16. Chiu, Y. L. & Greene, W. C. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu. Rev. Immunol.* **26,** 317–353 (2008).
17. Chen, H. *et al.* APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* **16,** 480–485 (2006).
18. Ostertag, E. M., Prak, E. T., DeBerardinis, R. J., Moran, J. V. & Kazazian, H. H. Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* **28,** 1418–1423 (2000).
19. Xie, Y., Rosser, J. M., Thompson, T. L., Boeke, J. D. & An, W. Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res.* **39,** e16 (2011).
20. Garcia-Perez, J. L. *et al.* LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16,** 1569–1577 (2007).
21. Wissing, S. *et al.* Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* **21,** 208–218 (2012).
22. De Fazio, S. *et al.* The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480,** 259–263 (2011).
23. Mathews, L. M., Chi, S. Y., Greenberg, N., Ovchinnikov, I. & Swergold, G. D. Large differences between LINE-1 amplification rates in the human and chimpanzee lineages. *Am. J. Hum. Genet.* **72,** 739–748 (2003).
24. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16,** 78–87 (2006).
25. Lee, J. *et al.* Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene* **390,** 18–27 (2007).
26. Mills, R. E. *et al.* Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.* **78,** 671–679 (2006).
27. Campbell, M. C. & Tishkoff, S. A. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* **9,** 403–433 (2008).
28. Gagneux, P. *et al.* Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl Acad. Sci. USA* **96,** 5077–5082 (1999).
29. Bowden, R. *et al.* Genomic tools for evolution and conservation in the chimpanzee: *Pan troglodytes ellioti* is a genetically distinct population. *PLoS Genet.* **8,** e1002504 (2012).
30. Beck, C. R. *et al.* LINE-1 retrotransposition activity in human genomes. *Cell* **141,** 1159–1170 (2010).

METHODS

Cell culture and retrovirus infection. Human ES cells HUES6 and H1, human iPS cell lines WT-33, ADRC-40 (human iPS cell lines 1 and 2 in this work, respectively) and WT-126 were previously described[31]. Fibroblasts from human GM22159 (WT-9), *P. troglodytes* (chimpanzees: PR00818 and PR01209) and *P. paniscus* (bonobos: AG05253 and PR01086) were from Coriell Cell Repositories (NJ) (Extended Data Table 1). All fibroblasts were cultured in MEM (Invitrogen) supplemented with 10% FBS (HyClone Laboratories). Retroviral vectors expressing *OCT4* (also known as *POU5F1*), *MYC*, *KLF4* and *SOX2* human cDNAs from Yamanaka's group[9] were obtained from Addgene. Recombinant viruses were produced by transient transfection in 293T cells, as previously described[32]. Two days after infection, cells were plated on mitotically inactivated mouse embryonic fibroblasts (Chemicon) with human ES cell medium. After 2–4 weeks, iPS cell colonies were picked manually and directly transferred to feeder-free conditions on matrigel-coated dishes (BD) using mTeSR1 (StemCell Technologies). Established iPS cell colonies were kept in feeder-free conditions indefinitely, and passed using mechanical dissociation. Embryoid-body-mediated differentiation in suspension was carried out for 10 days in the absence of growth factors. The use of chimpanzee and bonobo fibroblast samples was approved by the US Fish and Wildlife Service, under the permit MA206206. Protocols describing the use of iPS and human ES cells were previously approved by the University of California, San Diego (UCSD), the Salk Institute Institutional Review Board and the Embryonic Stem Cell Research Oversight Committee[31]. To generate stable shA3B cells lines, HUES6, WT-33 and ADRC-40 cells were transduced with lentiviruses expressing shRNAs and selected for puromycin resistance. pLKO.1-based lentiviral plasmids encoding shRNAs against *A3B* (RHS3979-99216651 and RHS3979-99216658) were obtained from Open Biosystems. Recombinant lentiviruses were produced by transient transfection on 293T cells as previously described[33].

Teratoma formation in nude mice. Around 1×10^6–3×10^6 cells were injected subcutaneously onto the dorsal flanks of nude mice (CByJ.Cg-Foxn1nu/J) anaesthetized with isoflurane. Five to six weeks after injection, teratomas were dissected, fixed overnight in 10% buffered formalin phosphate and embedded in paraffin. Tissues were then prepared for histopathologic analysis by the UCSD Mouse Phenotyping Services (http://mousepheno.ucsd.edu). In brief, the tissue was sectioned and stained with haematoxylin and eosin. Control mice injected with fibroblasts failed to form teratomas.

Karyotyping. Standard G-banding chromosome analysis was performed by Cell Line Genetics. Diploid human cells with $2n = 46$ chromosomes; bonobo and chimpanzee with $2n = 48$ chromosomes.

RNA extraction and RT–PCR. Total cellular RNA was extracted from ~5×10^6 cells using the RNeasy Protect Mini kit or RNeasy Plus kit (Qiagen), according to the manufacturer's instructions, and was reverse transcribed using the SuperScript III First-Strand Synthesis System RT–PCR from Invitrogen. For iPS cell markers, cDNA was amplified by PCR using Accuprime Taq DNA polymerase system (Invitrogen). Primer sequences are shown in Extended Data Table 1. PCR products were separated by electrophoresis on a 2% agarose gel, stained with ethidium bromide and visualized by ultraviolet illumination. Total RNA samples from human testis were obtained from Clontech. Small RNA was extracted using the mirVana kit (Ambion).

Quantitative RT–PCR. RNA was extracted using a QIAGEN RNeasy Plus kit or TRIzol (Life Technologies) and then retrotranscribed to cDNA with the Superscript III First-Strand synthesis system (Invitrogen). qRT–PCR reactions were carried out using SYBR Green mix (Roche) or TaqMan Assays (Life Technologies) using ABI Prism 7900HT sequence detection system (Applied Biosystems). The primers and Taqman sets used in this work are described in Extended Data Table 1. Data analysis was performed with SDS 2.3 software (Applied Biosystems). Primer efficiency was verified by linear regression to the standard curve. Values were normalized to *GAPDH*, *HPRT* or *ACTB*. Reactions were carried out in triplicate and data were analysed using the comparative ($\Delta\Delta C_t$) method. For *A3B* and *PIWIL2*, RNA levels were normalized to *GAPDH* or *ACTB* and represented as relative to iPS cell line 1. Relative *A3B* and *PIWIL2* mRNA levels normalized to *GAPDH* for each individual iPS cell line and fibroblasts are shown in Extended Data Fig. 9. The reduced levels of A3B in NHP iPS cell were not due to an A3B deletion polymorphism previously described in human individuals[34] (data not shown). For L1 RNA qRT–PCR, values representing the average of relative levels for L1 RNA (5′ UTR, ORF1 and ORF2) were calculated and normalized to actin mRNA levels. L1 levels are shown relative to iPS cell line 1. qRT–PCR analysis of L1-reporter expression in iPS cell lines transfected with L1–eGFP plasmid was carried out 60–72 h after transfection. At this time after transfection, eGFP RNA expressed from retrotransposed L1–eGFP will be insignificant compared to L1–eGFP plasmid-driven expression. eGFP levels were normalized to *GAPDH* or puromycin. L1–eGFP contains a puromycin expression cassette under *PGK* promoter control. Thus, puromycin expression can be used as normalizer for transfection.

iPS cells from two different individuals per species were transfected, and eGFP levels are shown relative to human iPS cells.

Plasmids. Human *A3B* cDNA from WT-33 and ADRC-40 iPS cells was amplified using Phusion high-fidelity polymerase (New England BioLabs), and primers are described in Extended Data Table 1. *A3B* cDNA fused to a haemagglutinin (HA) tag was then inserted into KpnI/XbaI-digested pcDNA3.1+ (pcDNA3-A3B) as previously described[35]. Similarly, *PIWIL2* cDNAs were amplified from human and NHP iPS cells as described above and inserted into pEF-BOS-EX using EcoRI/SalI (ref. 36). The plasmid expressing APOBEC3G was previously described[35].

RNA library generation and deep sequencing. PolyA$^+$ RNA was fragmented and prepared into sequencing libraries using the Illumina TruSeq RNA sample preparation kit and analysed on an Illumina HiSeq 2000 sequencer at the UCSD Biomedical Genomics Laboratory (BIOGEM). cDNA libraries were prepared from four human, two chimpanzee and two bonobo iPS cell lines derived from fibroblasts (two clones each, except for human WT-9 and WT-126), and two human ES cell lines (HUES6 and H1). Libraries were sequenced using paired-end $2 \times$ 100-bp (base pair) reads at a depth of 15–30 million reads per library (250 ± 25 bp (mean ± s.d.) fragments) (Extended Data Fig. 1b). Paired end reads from all libraries were mapped to both the human (hg19, GRCh37) and chimpanzee (panTro3, CGSC 2.1.3) genomes using STAR (v2.2.0c)[37]. To compare gene expression between human and NHP iPS cells, we first mapped paired end reads from all libraries to both human and chimpanzee genomes and then calculated gene expression read counts relative to human RefSeq transcripts. Owing to the lack of annotation in the chimpanzee genome, human gene models (RefSeq) were used to quantify gene expression. To avoid bias introduced by genome insertions and deletions, only reads mapping to both the human and chimpanzee genomes uniquely were used from each sample when comparing gene expression values (~4% of reads mapped to only one genome per sample). To calculate gene expression, read counts in the exons of RefSeq transcripts were calculated using HOMER[38]. Gene expression clustering was carried out using Gene Cluster 3.0 and visualized with Java Tree View[39,40]. EdgeR was used to identify differentially expressed genes comparing human samples with NHPs, and pairwise between bonobo, chimpanzee and human[41]. Functional enrichment analysis was restricted to differentially expressed coding genes with false discovery rates less than 5% and a fold change greater than twofold. We further restricted genes to contain an average of ten normalized reads across sample groups to remove genes with very low expression. Gene Ontology functional enrichment for biological processes (level 2) was carried out using DAVID[42], *Homo sapiens* whole genome was set as background.

Small RNA library generation and deep sequencing. Small RNA (15–40-nucleotide) libraries were prepared using the Illumina TruSeq Small RNA sample preparation kit and analysed on an Illumina HiSeq 2000 sequencer at the Beijing Genomics Institute. Libraries were sequenced using single-end reads at a depth of 15–25 million reads per library. Adaptor sequences were clipped from the 3′ end of each read and then aligned to the human (hg19, GRCh37) genome or to the L1Hs consensus sequence (Repbase[43]) using Bowtie2 (v4.1.2)[44]. Reads aligning to miRBase-defined microRNA transcripts were quantified using HOMER. Matches to previously identified human piRNAs were restricted to small RNAs with lengths between 26 and 33 nucleotides with 5′ ends within 2 nucleotides of previously identified piRNA 5′ ends based on piRNABank (http://pirnabank.ibab.ac.in/)[45].

L1 retrotransposition. Reporter L1 elements are tagged with a reporter gene (eGFP or firefly luciferase) such that only cells that complete a round of retrotransposition will express the reporter gene[46]. Three L1 reporter plasmids were used in this work. L1–eGFP[18,46] was previously described and was a gift from J. V. Moran. L1-luciferase-tagged plasmids (pYX014 and pYX017)[19] were obtained from W. An. In pYX014, L1 is regulated by its native promoter (5′ UTR) and, in pYX017, by the heterologous promoter CAG. pYX014 and pYX017 plasmids contain a *Renilla* luciferase expression cassette that allows for control of transfection efficiency. L1 assays in 293T cells were carried out as previously described[47]. 293T cells were transfected with L1 reporter plasmid together with control plasmid or plasmids expressing A3B, A3G or PIWIL2 using polyethylenimine (PEI). L1-luciferase retrotransposition was measured by quantification of luciferase activity using the Dual-Glo luciferase reporter assay (Promega) and normalized to *Renilla* luciferase. L1-luciferase inhibition in the presence of A3B or PIWIL2 was independent of the promoter driving L1 expression (data not shown). Inhibition levels of L1 retrotransposition by A3B and PIWIL2 were comparable between the three L1 reporter plasmids used in this study.

Plasmid transfections of iPS cells were performed by electroporation of L1–eGFP plasmid following the manufacturer's instructions (Lonza/Amaxa Nucleofactor, Kit V). The cells were then cultured under normal conditions for 10 days and the percentage of retrotransposition was measured by FACS of eGFP-positive cells. Electroporation efficiency of the L1–eGFP plasmid in human and NHP iPS cell was controlled by transfecting a cassette expressing eGFP and analysed by FACS after

48 h. Human and NHP iPS cell lines had similar transfection efficiency rates. To test the effect of A3B and PIWIL2 overexpression on L1 activity in NHP iPS cells, human *A3B* and human, chimpanzee or bonobo *PIWIL2* cDNAs were electroporated. All experiments were performed at least three times independently. L1 mobility assays are shown as relative value compared to control plasmid transfections or human iPS cell line 1 and represented as mean ± s.e.m. of at least three independent experiments.

Identification and cloning of a retrotransposition-competent chimpanzee L1. To clone an intact L1 and generate a chimpanzee L1–eGFP reporter plasmid, we followed a modification of the strategy previously described[48]. Intact L1Pt elements were identified in the chimpanzee genome (CSAC 2.1.4/panTro4, UCSC) through Blat and L1Xplorer analyses[49]. Among the identified intact full-length L1Pt elements, we amplified the L1 element located in chromosome 7:11771100–11777132 of the chimpanzee genome from 0.2 ng of genomic DNA extracted from chimpanzee iPS cell 1. Primers were designed to match unique sequences flanking 5′ and 3′ of the L1Pt and PCR reactions were performed using Phusion High-Fidelity polymerase (NEB). PCR product was sequenced to confirm intactness (Extended Data Fig. 5). A second PCR was performed using the first PCR product as template to introduce a NotI site upstream of the 5′ end of L1Pt. The second PCR product was digested with NotI/BstZ17I (New England Biolabs) and inserted into NotI/BstZ17I-digested pL1-eGFP replacing the human-L1 element to generate L1IN71 using the Rapid Ligation kit (Roche). L1IN71 contains a full-length L1Pt element tagged with the eGFP retrotransposition reporter cassette. Primers used for cloning L1IN71 are show in Extended Data Table 1.

L1 promoter activity. Human and chimp L1 promoters (L1 5′ UTR) were amplified by PCR from L1–eGFP and L1IN71 plasmids, and inserted into XhoI/HindIII digested pGL4.10 (Promega) upstream firefly luciferase cDNA (L1 5′ UTR plasmids). To quantify L1 promoter activity, L1 5′ UTR plasmids were co-transfected into human and NHP iPS cell lines along with a plasmid expressing *Renilla* luciferase. Seventy-two hours after transfection, luciferase activity was quantified and firefly luciferase signal was normalized to the *Renilla* luciferase signal. Results are shown as relative to human L1 5′ UTR activity in human iPS cells. Two iPS cell lines from different individuals (iPS cell 1 and 2) per species were transfected. Primers used for cloning L1 5′ UTRs are show in Extended Data Table 1.

Quantification of reference genome-encoded L1 insertions. Quantification of L1-derived genomic sequences was based on Repbase defined elements annotated by RepeatMasker (http://www.repeatmasker.org). L1 genomic positions for human (hg19, GRCh37) and chimpanzee (panTro3, CGSC 2.1.3) genomes were downloaded from the UCSC Genome Browser annotation database[50]. Owing to the large number of unfinished gaps in the chimpanzee genome assembly greater than 2 kilobases (kb) in size, only truncated L1 elements between 100 bp and 1 kb in length were considered in this analysis. Most of these represented the 3′ end of L1 elements. L1 elements were separated based on their annotation as L1HS, L1Pt, L1PA2, L1PA3 or L1PA4 and were plotted as a histogram relative to their divergence values, which indicates the fractions of nucleotides that are mutated relative to the consensus element for each family. To estimate the variability of L1 coverage across the genome, each genome was fragmented into 1 megabase (Mb) sections and then was randomly sampled in ten separate groups to calculate the standard deviation in number of L1 elements across different regions of the genome. A strong concentration of L1 elements in a few specific regions of the genome would result in a very high variance between groups, whereas uniform insertion across the genome would result in a low variance. This standard deviation between each sampling was reported as a function of divergence for each class of L1 elements.

To identify reference L1 elements that were inserted into the genome after the last common ancestor for human and chimpanzee, L1 elements were mapped between homologous regions of each genome using the UCSC LiftOver tool. If an element failed to map between genomes, the 100 bp regions immediately upstream and downstream of the L1 element were also mapped between genomes using the LiftOver tool. If the upstream and downstream regions both mapped to the other genome, then the L1 element was mostly likely a result of a recent insertion. If only one or neither of the upstream and downstream regions mapped between genomes, the region was more likely to be the result of a genomic duplication or deletion and was discarded from the analysis. Error bars (s.d.) represent the differences in L1 density based on the sampling of different genomic regions and represents the variability of L1 coverage across the genomes.

Immunocytochemistry. Cells were fixed in 4% paraformaldehyde and then permeabilized with 0.5% Triton X-100 in PBS. Cells were then blocked in 5% donkey serum for 1 h before incubation with primary antibody overnight at 4 °C. After three washes with PBS, cells were incubated with secondary antibodies conjugated to fluorophors (Jackson Immuno Research) for 1 h at room temperature. Fluorescence was detected using a Zeiss inverted microscope.

Immunoblotting. Immunoblotting was performed as previously described[35]. Cell pellets were lysed in lysis buffer supplemented with Complete protease inhibitor cocktail (Roche) for 30 min on ice[35]. Protein concentrations from whole cell lysates were quantified by BCA assay (Bio-Rad). Proteins were separated in 4–12% Acrylamide Bis-Tris NuPage gels in MOPS buffer (Invitrogen) and transferred onto Hybond nitrocellulose membranes (Amersham Biosciences).

Antibodies. Primary antibodies used in this study were: Tra-1-81 (1:100, Millipore, MAB4381), Nanog (1:500, R&D Systems, AF1997), APOBEC3B (D-15) (1:500, Santa Cruz, sc-86289), PIWIL2 (1:1,000, R&D Systems, AF6558), GFP (1:200, Molecular Probes-Invitrogen, A-6455), Flag (1:1,000 Sigma, F7425) and HA (1:1,000 Covance, MMS-101R). All secondary antibodies were purchased from Jackson ImmunoResearch.

PIWIL2 RNPs immunoprecipitation and end labelling. Tetracycline-inducible human iPS cells expressing flag-tagged PIWIL2 were generated by transduction with lentiviruses (Lv)[51]. Cells were first transduced with an Lv-expressing tetracycline transactivator rtTA (LvXEtO). After 10 days of culture in growth media with neomycin (neo), neo-resistant colonies were then transduced with a lentivirus expressing Flag–PIWIL2 under the control of a tetracycline-inducble promoter (LvXTP-FlagPIWIL2) and selected for resistance to puromycin. For PIWIL2 RNP immunoprecipitation, ~3 × 10^7 human iPS cells were treated with doxycycline for 72 h, and pelleted cells were resuspended in 1 ml lysis buffer 1 (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM MgCl$_2$, 0.5% NP40, 1% glycerol, 1 mM dithiothreitol (DTT), 0.1 U μl^{-1} RNase inhibitor (Ambion), Complete EDTA-free protease inhibitor (Roche)). Cell lysates were cleared by centrifugation at 20,000*g* for 20 min at 4 °C. Cleared lysates were incubated with EZview Red FLAG M2 Affinity Gel (Sigma) for 3 h at 4 °C and washed five times with wash buffer (lysis buffer 1 without glycerol). Co-immunoprecipitated RNAs were extracted with Trizol, followed by precipitation with isopropanol and glycogen (Ambion). Isolated RNA was 5′ labelled with [γ-^{32}P]ATP using T4 polynucleotide kinase (NEB), resolved on 15% PAGE TBE urea gels along with radiolabelled Decade size marker (Ambion) and visualized in a Typhoon phosphorimager (Amersham Biosciences). Control immunoprecipitations were carried out with lysates from cells without doxycycline induction, from doxycycline-induced eGFP-expressing human iPS cells or with control antibody (anti-HA, Roche, 3F10).

Data deposition. RNA-seq and small RNA-seq data have been deposited in the GEO under accession number GSE47626. GenBank accession numbers: KF651164 (*P. paniscus* PIWIL2), KF651165: (*H. sapiens* PIWIL2), KF651166 (*P. troglodytes* PIWIL2), KF651167 (*H. sapiens* APOBEC3B), KF651168 (*P. troglodytes* APOBEC3B), KF651169 (*P. paniscus* APOBEC3B) and KF661301 (L1Pt in chimp-L1 plasmid).

31. Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143,** 527–539 (2010).
32. Muotri, A. R., Nakashima, K., Toni, N., Sandler, V. M. & Gage, F. H. Development of functional human embryonic stem cell-derived neurons in mouse brain. *Proc. Natl Acad. Sci. USA* **102,** 18644–18648 (2005).
33. Landry, S., Narvaiza, I., Linfesty, D. C. & Weitzman, M. D. APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO Rep.* **12,** 444–450 (2011).
34. Kidd, J. M., Newman, T. L., Tuzun, E., Kaul, R. & Eichler, E. E. Population stratification of a common *APOBEC* gene deletion polymorphism. *PLoS Genet.* **3,** e6 (2007).
35. Narvaiza, I. *et al.* Deaminase-independent inhibition of parvoviruses by the APOBEC3A cytidine deaminase. *PLoS Pathog.* **5,** e1000439 (2009).
36. Mizushima, S. & Nagata, S. pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Res.* **18,** 5322 (1990).
37. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
38. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
39. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–1454 (2004).
40. Saldanha, A. J. Java Treeview–extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).
41. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
42. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4,** P3 (2003).
43. Jurka, J. *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110,** 462–467 (2005).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
45. Sai Lakshmi, S. & Agrawal, S. piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* **36,** D173–D177 (2008).
46. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87,** 917–927 (1996).
47. Bulliard, Y. *et al.* Structure–function analyses point to a polynucleotide-accommodating groove essential for APOBEC3A restriction activities. *J. Virol.* **85,** 1765–1776 (2011).

48. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA* **100,** 5280–5285 (2003).

49. Penzkofer, T., Dandekar, T. & Zemojtel, T. L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res.* **33,** D498–D500 (2005).

50. Fujita, P. A. *et al.* The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* **39,** D876–D882 (2011).

51. Ladewig, J. *et al.* Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nature Methods* **9,** 575–578 (2012).

a

Species	Sex	Name/Source
Homo sapiens	Female	hES HUES6 / (HSCI) Embryonic Stem Cell
Homo sapiens	Male	hES H1 / (WiCell) Embryonic Stem Cell
Homo sapiens	Female	WT-33 (iPS1*) / Fibroblast (Marchetto et. *al* Cell 2010)
Homo sapiens	Male	ADRC-40 (iPS2*) / Fibroblast (Marchetto et. *al* Cell 2010)
Homo sapiens	Male	GM22159 (WT-9* iPS) / Fibroblast (Coriell Cell Repositories)
Homo sapiens	Male	WT-126 (iPS) / Fibroblast (Marchetto et. *al* Cell 2010)
Pan paniscus (Bonobo, pigmy chimp)	Male	PR01086 (iPS1*) / Fibroblast (Coriell Cell Repositories)
Pan paniscus (Bonobo, pigmy chimp)	Male	AG05253 (iPS2*) / Fibroblast (Coriell Cell Repositories)
Pan troglodytes (Chimpanzee)	Male	PR01209 (iPS1*) / Fibroblast (Coriell Cell Repositories)
Pan troglodytes (Chimpanzee)	Female	PR00818 (iPS2*) / Fibroblast (Coriell Cell Repositories)

* nomenclature used in this study

b

RNA-Seq Samples	Mapped Read Totals
Human WT-33 iPS1 A	25430637
Human WT-33 iPS1 B	13839658
Human ADRC-40 iPS2 A	30337397
Human ADRC-40 iPS2 B	22682712
Human WT-126 iPS	28193749
Human GM22159 (WT-9) iPS	25613490
Bonobo PR01086 iPS1 A	13137905
Bonobo PR01086 iPS1 B	15666314
Bonobo AG05253 iPS2 A	11520455
Bonobo AG05253 iPS2 B	62143394
Chimp PR01209 iPS1 A	22705535
Chimp PR01209 iPS1 B	10917561
Chimp PR00818 iPS2 A	24052999
Chimp PR00818 iPS2 B	28197740
Human ES HUES6	26114155
Human ES H1	29868106
Human NPC (derived from HUES6)	26814443

c

Gene Ontology Term	Count	PValue	Benjamini
0007154:cell communication	91	4.72×10^{-8}	9.39×10^{-6}
0007610:behavior	60	3.61×10^{-7}	3.59×10^{-5}
0007155:cell adhesion	80	3.81×10^{-7}	2.52×10^{-5}
0042221:response to chemical stimulus	124	1.73×10^{-6}	8.61×10^{-5}
0051239:regulation of multicellular organismal process	94	9.70×10^{-6}	3.86×10^{-4}
0009605:response to external stimulus	89	5.34×10^{-5}	1.77×10^{-3}
0007275:multicellular organismal development	230	5.70×10^{-5}	1.62×10^{-3}
0032879:regulation of localization	64	8.96×10^{-5}	2.23×10^{-3}
0048856:anatomical structure development	202	2.55×10^{-4}	4.60×10^{-3}
0065008:regulation of biological quality	126	3.67×10^{-4}	6.08×10^{-3}

d

Gene Ontology Term	Count	PValue	Benjamini
0006323:DNA packaging	13	5.89×10^{-3}	6.76×10^{-1}
0007586:digestion	11	7.14×10^{-3}	4.96×10^{-1}
0008037:cell recognition	8	1.06×10^{-2}	4.94×10^{-1}
0022414:reproductive process	48	1.44×10^{-2}	5.00×10^{-1}
0006950:response to stress	94	1.58×10^{-2}	4.56×10^{-1}
0009605:response to external stimulus	54	2.91×10^{-2}	6.10×10^{-1}
0051093:negative regulation of developmental process	19	4.14×10^{-2}	6.84×10^{-1}
0009791:post-embryonic development	8	4.34×10^{-2}	6.53×10^{-1}
0042445:hormone metabolic process	10	4.67×10^{-2}	6.37×10^{-1}
0050878:regulation of body fluid levels	12	5.02×10^{-2}	6.26×10^{-1}

Extended Data Figure 1 | Cell lines used, number of mapped reads per sample in RNA-seq and gene ontology enrichment analysis for differentially expressed genes. a, Origin of iPS cells used or generated in this study. **b,** Total number of mapped reads per sample in RNA-seq. **c, d,** Gene ontology (GO) enrichment analysis of differentially expressed genes. **c,** Top 10 enriched GO terms for genes with higher expression in human versus NHP iPS cells. **d,** Top 10 enriched GO terms for genes highly expressed in NHP versus human iPS cells. GO analysis was restricted to differentially expressed protein-coding genes (FDR < 0.05 and fold change > 2). GO enrichment for biological processes (level 2) was performed using DAVID. Figure shows GO term, number of genes (count), and *P* values for EASE score and Benjamini adjustment.

a

APOBEC3B

```
                    20                  40                  60                  80                  100
Human   MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFEPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC
Chimp   MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDC
Bonobo  MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMYFLSWFCGNQLSAYKCFQITWFVSWTPCPDC

                    120                 140                 160                 180                 200
Human   VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTF
Chimp   VAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTF
Bonobo  VAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDYEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTF

                    220                 240                 260                 280                 300
Human   NFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQEN
Chimp   NFNNDPLVLRRHQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQEN
Bonobo  NFNNDPLVLRRHQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQEN

                    320                 340                 360                 380
Human   THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN
Chimp   THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN
Bonobo  THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN
```

b

PIWIL2

```
                    20                  40                  60                  80                  100
Human   MDPFRPSFRGQSPIHPSQCQAVRMPGCWPQASKPLDPALGRGAPAGRGHVFGKPEEPSTQRGPAQRESVGLVSMFRGLGIETVSKTPLKREMLPSGRGIL
Chimp   MDPFRPSFRGQSPIHPSQCQAVRMPGCWPQASKPLDPALGRGAPAGRGHVFGKPEEPSTQRGPAQRESVGLVSMFQGLGIETVSKTPLKREMLPSGRGIL
Bonobo  MDPFRPSFRGQSPIHPSQCQAVRMPGCWPQASKPLDPALGRGAPAGRGHVFGKPEEPSTQRGPAQRESVGLVSMFQGLGIETVSKTPLKREMLPSGRGIL

                    120                 140                 160                 180                 200
Human   GRGLSANLVRKDREELSPTFWDPKVLAAGDSKMAETSVGWSRTLGRGSSDASLLPLGRAAGGISREVDKPPCTFSTPSRGPPQLSSPPALPQSPLHSPDR
Chimp   GRGLSANLVRKDREELSPTFWDPKVLAAGDSKMAEASVGWSRTLGRGSSDASLLPLGRAAGGIGRGVDKPPCTFSTLSRGPTQLSSSPALPQSPLHSPDR
Bonobo  GRGLSANLVRKDREELSPTFWDPKVLAAGDSKMAEASVGWSRTLGRGSSDASLLPLGRAAGGIGRGVDKPPCTFSTLSRGPTQLSSPPALPQSPLHSPDR

                    220                 240                 260                 280                 300
Human   PLVLTVEHKEKELIVKQGSKGTPQSLGLNLVKIQCHNEAVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGSILYLPVKLQQVLELKSQRKTDSA
Chimp   PLVLTVEHKEKELIVKQGSKGTPQSLGLNLVKIQCHNEAVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGSILYLPVKLQQVLELKSQRKTDSA
Bonobo  PLVLTVEHKEKELIVKQGSKGTPQSLGLNLVKIQCHNEAVYQYHVTFSPNVECKSMRFGMLKDHQAVTGNVTAFDGSILYLPVKLQQVLELKSQRKTDSA

                    320                 340                 360                 380                 400
Human   EISIKIQMTKILEPCSDLCIPFYNVVFRRVMKLLDMKLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDCVLDVMHAIYQQ
Chimp   EISIKIQMTKILEPCSDLCIPFYNVVFRRVMKLLDMKLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDCVLDVMHAIYQQ
Bonobo  EISIKIQMTKILEPCSDLCIPFYNVVFRRVMKLLDMKLVGRNFYDPTSAMVLQQHRLQIWPGYAASIRRTDGGLFLLADVSHKVIRNDCVLDVMHAIYQQ

                    420                 440                 460                 480                 500
Human   NKEHFQDECTKLLVGNIVITRYNNRTYRIDDVDWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQDNHGMLLKGEILLLPELSFM
Chimp   NKEHFQEECTKLLVGNIVITRYNNRTYRIDDVDWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQNNHGMLLKGEILLLPELSFM
Bonobo  NKEHFQEECTKLLVGNIVITRYNNRTYRIDDVDWNKTPKDSFTMSDGKEITFLEYYSKNYGITVKEEDQPLLIHRPSERQNNHGMLLKGEILLLPELSFM

                    520                 540                 560                 580                 600
Human   TGIPEKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIEGRVLPMERINLKNTSFITSQELNWVKEVTRDPSILT
Chimp   TGIPEKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIEGRVLPMERINLKNTSFITSQELNWVKEVTRDPSILT
Bonobo  TGIPEKMKKDFRAMKDLAQQINLSPKQHHSALECLLQRIAKNEAATNELMRWGLRLQKDVHKIEGRVLPMERINLKNTSFITSQELNWVKEVTRDPSILT

                    620                 640                 660                 680                 700
Human   IPMHFWALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMGPRDDLYGAIKKLCCVQSPVPSQVVN
Chimp   IPMHFWALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMAPRDDLYGAIKKLCCVQSPVPSQVVN
Bonobo  IPMHFWALFYPKRAMDQARELVNMLEKIAGPIGMRMSPPAWVELKDDRIETYVRTIQSTLGAEGKIQMVVCIIMAPRDDLYGAIKKLCCVQSPVPSQVVN

                    720                 740                 760                 780                 800
Human   VRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFY
Chimp   VRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFY
Bonobo  VRTIGQPTRLRSVAQKILLQINCKLGGELWGVDIPLKQLMVIGMDVYHDPSRGMRSVVGFVASINLTLTKWYSRVVFQMPHQEIVDSLKLCLVGSLKKFY

                    820                 840                 860                 880                 900
Human   EVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAFENYQPKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGC
Chimp   EVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAFENYQPKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGC
Bonobo  EVNHCLPEKIVVYRDGVSDGQLKTVANYEIPQLQKCFEAFENYQPKMVVFVVQKKISTNLYLAAPQNFVTPTPGTVVDHTITSCEWVDFYLLAHHVRQGC

                    920                 940                 960
Human   GIPTHYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL
Chimp   GIPTHYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL
Bonobo  GIPTHYVCVLNTANLSPDHMQRLTFKLCHMYWNWPGTIRVPAPCKYAHKLAFLSGHILHHEPAIQLCENLFFL
```

Extended Data Figure 2 | Amino acid alignment of A3B and PIWIL2.
a, **b**, Protein sequences of human, chimp and bonobo A3B (**a**) or PIWIL2
(**b**) were aligned using ClustalW. **a**, Alignment of A3B showing >93% identity
between human and NHP proteins. **b**, Alignment of PIWIL2 showing >98%
identity between human and NHP proteins.
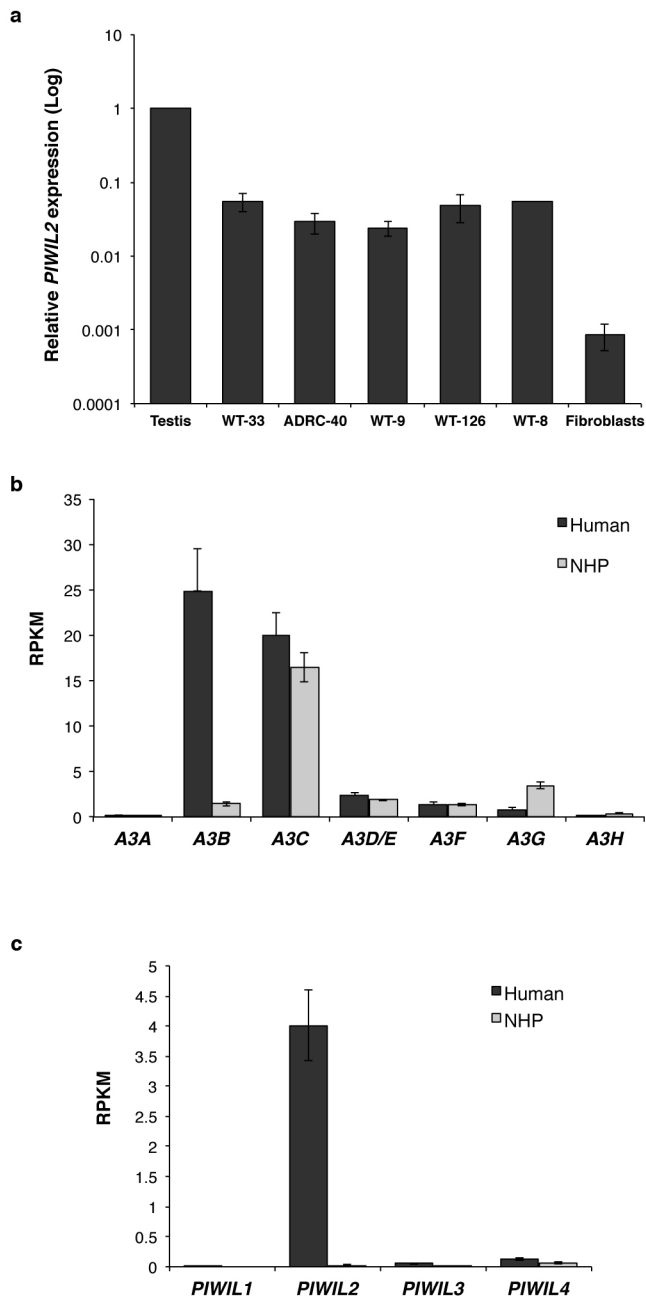
Extended Data Figure 3 | mRNA levels of *APOBEC3* and PIWI-like protein family members in iPS cells. a, Comparative analysis of *PIWIL2* mRNA levels. qRT–PCR analysis of *PIWIL2* mRNA levels in human testis, human iPS cell lines, and available fibroblasts from which the iPS cell lines were derived. mRNA levels were normalized to *GAPDH* and shown relative to human testis (mean ± s.e.m.; $n = 3$ biological replicates). Compared to testis, *PIWIL2* levels are 20–40 fold lower in iPS cells and ~1,100-fold lower in fibroblasts.
b, c, Quantification of mRNA levels of *APOBEC3* and PIWI-like family members in human and NHP iPS cells by RNA-seq. Increased mRNA levels in human iPS cells are restricted for *APOBEC3B* and *PIWIL2*. *y* axes in **b** and **c** denote the reads per kilobase per million mapped reads (RPKM).
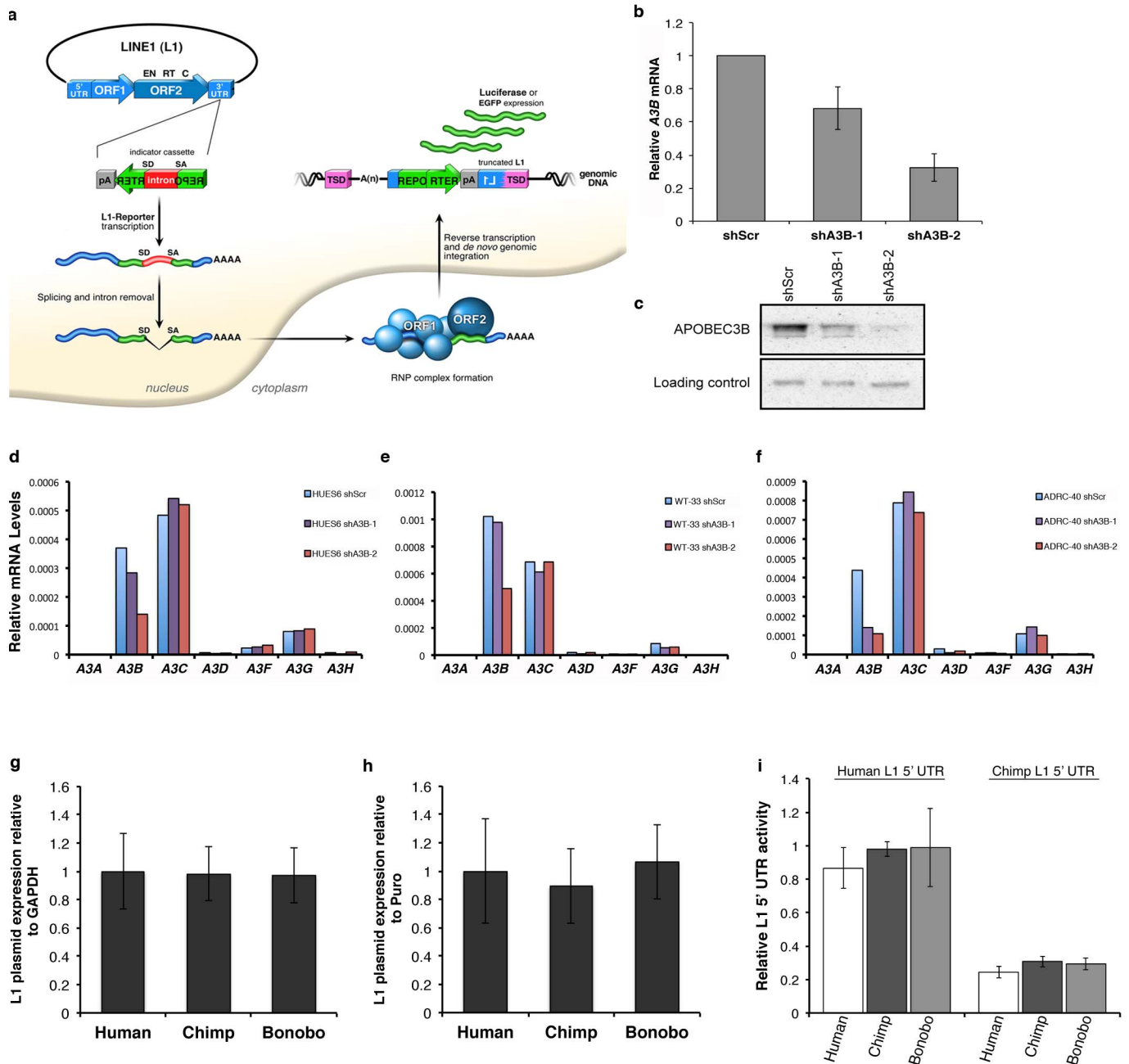
Extended Data Figure 4 | L1 reporter activity in iPS cells. **a**, L1 retrotransposition reporter system. The L1-reporter plasmid contains a retrotransposition-competent human L1 element and carries either an eGFP or a luciferase reporter construct in its 3′ UTR region. The reporter gene is interrupted by an intron in the same transcriptional orientation as the L1 transcript. This arrangement ensures that eGFP/luciferase-positive cells will arise only when a transcript initiated from the promoter driving L1 expression is spliced, reverse transcribed, and integrated into chromosomal DNA, thereby allowing expression of the reporter gene from a heterologous promoter. **b–f**, Efficient A3B knockdown in human ES and iPS cells. **b**, Stable shRNA-mediated knockdown of A3B in human ES cells (HUES6) using lentivirus expressing different shRNAs against *A3B* (shA3B-1 and shA3B-2) or scrambled control (shScr). Levels of *A3B* expression were normalized to *GAPDH* and shown relative to shScr (mean ± s.e.m.; *n* = 3 biological replicates). **c**, Western blot confirming stable A3B knockdown in human ES cells. **d–f**, shRNA-mediated knockdown in human ES cells (HUES6) and iPS cell lines 1 and 2 (WT-33 and ADRC-40 , respectively) was specific for *A3B*.

g–h, qRT–PCR analysis of plasmid expression in iPS cell lines transfected with L1–eGFP plasmid. Total RNA samples were obtained 60–72 h after transfection. L1 plasmid expression was normalized to *GAPDH* (**g**) or puromycin (**h**). L1–eGFP contains a puromycin expression cassette under *PGK* promoter control. Thus, puromycin expression was used as normalizer for transfection. iPS cells from two different individuals per species were transfected, and eGFP levels are shown as relative to human iPS cells. No significant differences were observed for L1 plasmid expression between human and NHP iPS cell lines (mean ± s.e.m.; *n* = 3 biological replicates). **i**, Relative L1 5′ UTR promoter activity. Human and chimp L1 promoters (L1 5′ UTR) controlling firefly luciferase were transfected into human and NHP iPS cell lines. *Renilla* luciferase was co-transfected as control. Luciferase activity was quantified as firefly luciferase units relative to *Renilla* luciferase units. Results are shown as normalized to human L1 5′ UTR activity in human iPS cell. iPS cells from two different individuals per species were transfected. No significant differences were observed for L1 promoter activities between human and NHP iPS cell lines (mean ± s.e.m.; *n* = 4 biological replicates).

Extended Data Figure 5 | Nucleic acid alignment of human and chimpanzee L1 elements. Sequence of the chimpanzee L1Pt element cloned and used to generate the chimpanzee L1–eGFP tagged reporter plasmid (L1IN71) (top sequence). *LRE3*: human L1 (bottom sequence).

a

DOX-inducible Flag-PIWIL2

- + - - + + Dox

input   HA Flag HA Flag   IP-Antibody

← PIWIL2

b

DOX-inducible Flag-PIWIL2

αHA   αFlag   IP antibody

30 nt

20 nt

piRNAS

d

Number of mapped reads and piRNAs mapped per sample for small RNA-seq

Sample	Total Mapped Reads
Human iPS1	17787803
Human iPS2	20057883

Sample	Total 26-33 nt Reads*	Unique positions
Human iPS1	5332430	311300
Human iPS2	4610072	200360

Sample	Total 26-33 nt Reads mapping to L1Hs*	Unique positions to L1Hs
Human iPS1	2348	722
Human iPS2	2453	716

* 2 mismatches allowed.

e

Annotated piRNAs (piRNABank) mapping to the human genome detected by RNA-seq:
hiPS1: 272
hiPS2: 229
(181 in common)

Annotated piRNAs (piRNABank) mapping to consensus L1Hs detected by RNA-seq:
hiPS1: 12
hipS2: 10
(7 in common)

c

f

g

LINE1 complementarity	Position on L1Hs	strand
annotated piRNA sequence	piRNA	Hits
read sequence	Seq. tag number	#reads

hiPS1

L1HS	64	-
TCGCTCACGCTGGGGAGCTGTAGACCGGAGC	hsa_piR_005007	874
--GCTCACGCTGGGGAGCCGTAGACCGGAG	t490804-1	2

L1HS	64	-
-TCGCTCACGCTGGGGAGCTGTAGACCGGAGC	hsa_piR_005007	874
GTCGCTCACGCTGGGGAGCTGTAGACCGGAG	t464559-1	2

L1HS	64	-
-TCGCTCACGCTGGGGAGCTGTAGACCGGAGC	hsa_piR_005007	874
GTCGCTCACGCTGGGGAGCTGTAGACCGGGG	t708895-0	1

L1HS	64	-
-TCGCTCACGCTGGGGAGCTGTAGACCGGAGC	hsa_piR_005007	874
GTTGCTCACGCTGGGGAGCTGTAGACCGGAG	t694268-0	1

L1HS	394	-
TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
-GCTAGCAATCAGCGAGACTCCGTGGGCG	t082439-1	7

L1HS	394	-
TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
-GCTAGCAATCAGCGAGACTCCGTGGG	t491030-0	2

L1HS	394	-
TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
-GCTAGCAATCAGCGAGACTCCGTGGGCGTA	t321643-1	3

L1HS	887	-
TGGTCTTTGATGATGGTGATGTACAGA	hsa_piR_014658	2842
--GTATTTGATGATGGTGATGTACAGATGGG	t199191-3	4

L1HS	1378	+
TCTACGTCTGATTGATGTACCTGAAAGTGA	hsa_piR_005263	21
-CTACGTCTGACTGGTGTACCTGAAAGTGATG	t523911-0	2

L1HS	1379	+
CTACATCTGATTGGTGTACCTGAAAGTGA	hsa_piR_017002	827
CTACGTCTGACTGGTGTACCTGAAAGTGATG	t523911-0	2

L1HS	2075	+
TGGATAAAGAGTCAAGACCCGTCAGTGTGC	hsa_piR_012728	29
TGGATAAAGAGTCAAGACCCATCAGTGTGCTG	t275379-2	3

L1HS	3224	-
TCTGATGGTAGTTTGTGTTTCTGTGGG	hsa_piR_005691	121
TCTGATGGTAGTTCGTGATTTCTGTGGGATCGGT	t436256-0	2

L1HS	5067	+
TCAGAGTGAACAGGCAACCTACAAAATGG	hsa_piR_002528	4397
-CAGAGTGAACAGGCAACCTACAACATGGG	t542528-0	2

L1HS	5254	-
--TTGGCTGCATAGATGTCTTCTTTTGAGAAGT	hsa_piR_018467	229
GTTGGCTGCATAAATGTCTTCTTTTGAGAAGTG	t296674-2	3

L1HS	5254	-
--TTGGCTGCATAGATGTCTTCTTTTGAGAAGT	hsa_piR_018467	229
TTTTGGCTGCATAAATGTCTTCGTTTGAGAA	t172031-0	4

L1HS	5296	-
TCTCTGATGGCCAGTGATGATGAACGTTTT	hsa_piR_005541	2
--TCTGATGGCCAGTGATGATGAGCATTTTT	t436262-0	2

L1HS	5296	-
-TCTCTGATGGCCAGTGATGATGAACGTTTT	hsa_piR_005541	2
TTCTCTGATGGCCAGTGATGATGAGCAT	t620927-0	1

L1HS	5296	-
-TCTCTGATGGCCAGTGATGATGAACGTTTT	hsa_piR_005541	2
TTCTCTGATGGCCAGTGATGATGAGCATTT	t122731-3	5

L1HS	5634	-
--TTGATGGACATTTGGGTTGGTTCCAAGTC	hsa_piR_018145	3262
CATTTTTGGACATTTGGGTTGGTTCCAAGTC	t229813-2	4

L1HS	5802	+
TAGGTGGGAATTGAACAATGAGATCA	hsa_piR_023244	3911
--GGTGGGAATTGAACAATGAGAACACATGGA	t470098-0	2

L1HS	5802	+
TAGGTGGGAATTGAACAATGAGATCA	hsa_piR_023244	3911
--GGTGGGAATTGAACAATGAGAGCACAT	t201215-1	4

hiPS2

L1HS	45	-
TAGACCGGAGCTGTTCCTGTTCGGCCATC	hsa_piR_022583	34
--GACCCGGAGCTGTTCCTATTCGGCCATCTT	t369109-0	2

L1HS	45	-
-TAGACCGGAGCTGTTCCTGTTCGGCCATC	hsa_piR_022583	34
GTTGACTGGAGCTGTTCCTATTCGGCCAT	t220289-2	3

L1HS	64	-
TCGCTCACGCTGGGGAGCTGTAGACGGGAGC	hsa_piR_005007	874
--CGCTCACGCTGGGGAGCTGTAGACCGGA	t582539-0	1

L1HS	394	-
TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
-GCTAGCAATCAGCGAGATTCCGTGGGC	t107811-4	5

L1HS	394	-
TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
TGCTTAGCAATCAGCGAGACTCCGTGGGCG	t311793-1	2

L1HS	394	-
--TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
TGTGCTAGCAATCAGCGAGACTCCGTGGGCG	t094873-0	5

L1HS	394	-
--TGCTAGCAATCAGCGAGATTCCGTGGGC	hsa_piR_011495	438
TGTGCTAGCAATCAGCGAGATTCCGTGGGC	t133798-2	4

L1HS	887	-
TGGTCTTTGATGATGGTGATGTACAGA	hsa_piR_014658	2842
--GGTCTTTGATGATGGTGAACAGATGGGTT	t517608-0	1

L1HS	1379	+
CTACATCTGATTGGTGTACCTGAAAGTGA	hsa_piR_017002	827
--ACGTCTGATTGGTGTACCTGAAAGTGACGG	t647191-0	1

L1HS	2075	+
TGGATAAAGAGTCAAGACCCGTCAGTGTGC	hsa_piR_012728	29
TGGATAAAGAGTCAAGACCCATCAGTG	t096174-3	5

L1HS	3224	-
TCTGATGGTAGTTTGTGTTTCTGTGGG	hsa_piR_005691	121
--TGATGGTAGTTTGTATTTCTGTGGGATCG	t313216-0	2

L1HS	3473	-
TAGTTTCAGAAGGAATGGTACCAGCTCC	hsa_piR_001450	3792
TAGTTTCAGAAGGAATGGTACCAGCTCCT	t215966-1	3

L1HS	5067	+
--TCAGAGTGAACAGGCAACCTACAAAATGG	hsa_piR_002528	4397
CATCAGAGTGAACAGGCAACCTATAAAATGG	t393412-1	2

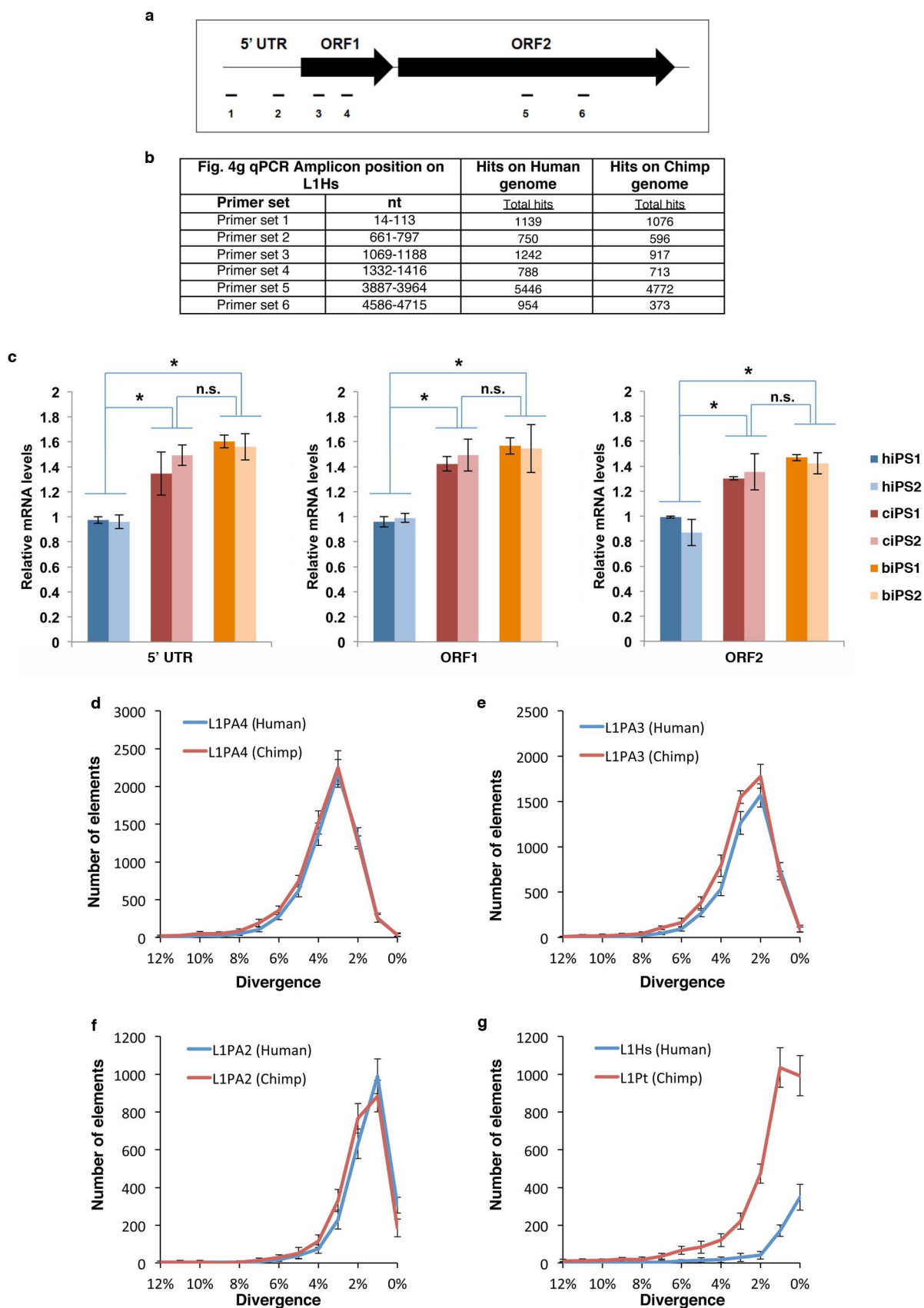| L1HS | 5518 | - |
| AGTAATGGGATTGCTGGGTCAAATGGTA | hsa_piR_001104 | 4409 |
| -GTAATGGGATTGCTGGGTCAAATGGTA | t340964-1 | 2 |

Extended Data Figure 6 | Immunoprecipitation of piRNAs associated with PIWIL2 in human iPS cells and annotated piRNAs mapping to consensus L1Hs in iPS cells. **a**, Immunoprecipitation of PIWIL2 RNPs using Flag-tag antibodies from Tet-inducible Flag-tagged PIWIL2 human iPS cells after addition of doxyclycine to the culture media. HA-tag antibody was used as control. **b**, [γ-^{32}P]ATP end-labelling of RNAs associated with Flag–PIWIL2 RNPs. Signal in the piRNAs size range is detected only in anti-Flag but not in control antibody anti-HA immunoprecipitates. **c**, Size distribution of RNA reads detected by small RNA-seq from small RNAs samples extracted from human iPS cell lines. **d**, Number of mapped reads per sample in small RNA-seq. **e**, Number of annotated piRNAs (piRNAbank) detected by RNA-seq in human iPS cells 1 and 2. **f**, Characterization of 5′ end of piRNAs detected in human iPS cells relative to annotated piRNAs. Read count distribution relative to piRNA 5′ ends (piRNAbank). **g**, Sequences of annotated piRNAs (piRNAbank) mapping to consensus L1Hs detected in human iPS cells 1 and 2. The 26–33-nucleotide RNA reads from human iPS cell lines 1 and 2 characterized by RNA-seq are aligned to annotated piRNAs mapping to the consensus L1Hs sequence. Analysis of mapping sequences was performed allowing two mismatches.
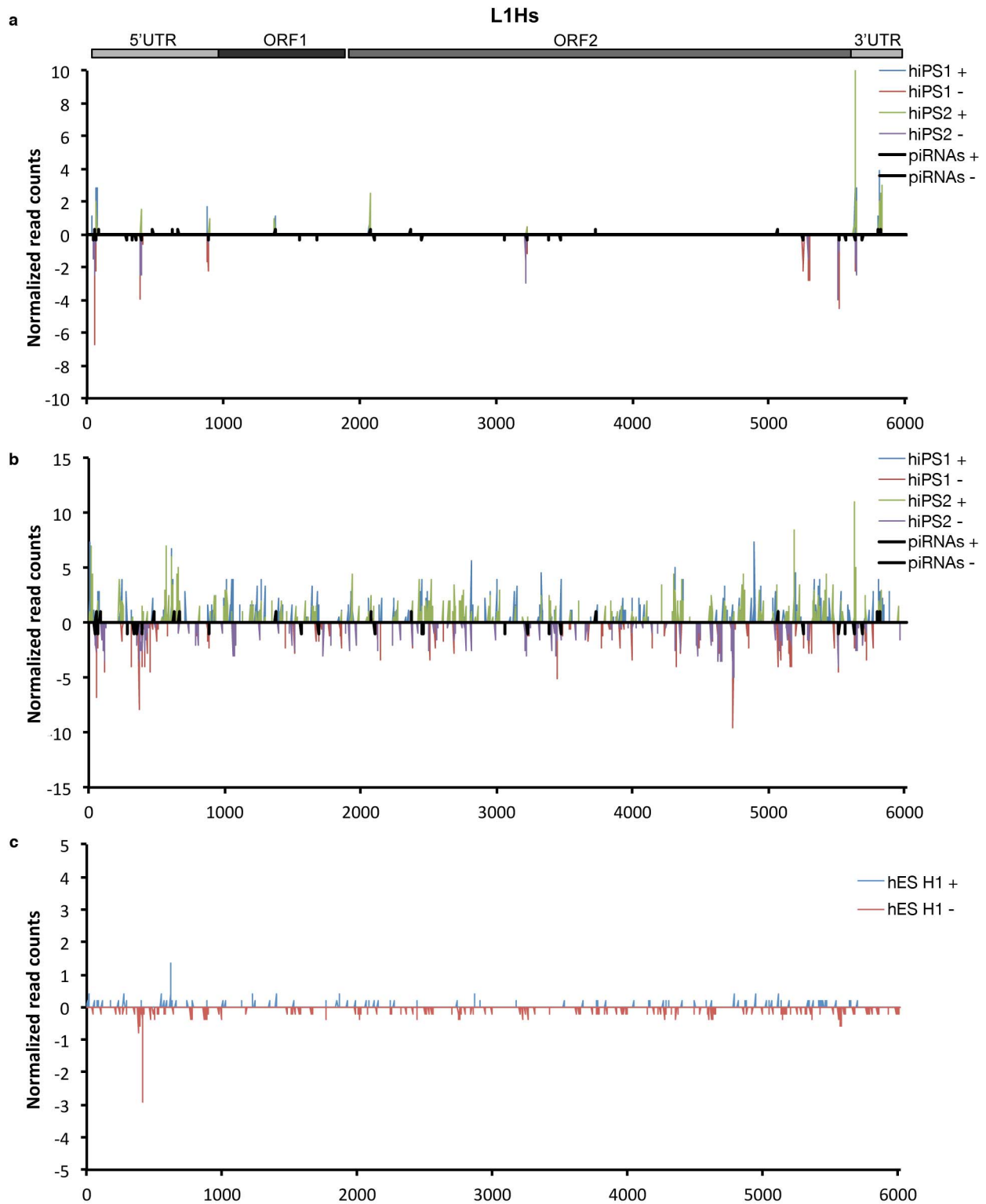
Extended Data Figure 7 | Mapping of 26–33-nucleotide RNAs in human iPS cells to consensus L1Hs. **a**, Mapping of annotated piRNAs (piRNAbank) detected by RNA-seq from human iPS cell lines to the consensus sequence for L1Hs (from Repbase). All annotated piRNAs (piRNAbank) complementary to L1Hs are indicated (black bars). **b**, Total 26–33-nucleotide RNA reads characterized by small RNA-seq mapped to L1Hs. **c**, Similar analysis as in **b** of ENCODE data for small RNAs from H1 cells. Positive and negative values indicate sense (+) and antisense (−) piRNAs, respectively. Schematic representation of the L1Hs element is shown (top). *y* axes represent read counts normalized to 10^7 reads per experiment.

a

b

Fig. 4g qPCR Amplicon position on L1Hs		Hits on Human genome	Hits on Chimp genome
Primer set	nt	Total hits	Total hits
Primer set 1	14-113	1139	1076
Primer set 2	661-797	750	596
Primer set 3	1069-1188	1242	917
Primer set 4	1332-1416	788	713
Primer set 5	3887-3964	5446	4772
Primer set 6	4586-4715	954	373

c

d

e

f

g

Extended Data Figure 8 | Higher levels of endogenous L1 RNA and recent species-specific L1 elements in chimpanzee. a, Scheme of amplicons mapped to the L1Hs consensus sequence. Six primer pairs (two per region) were used for quantification of 5′ UTR, ORF1 and ORF2. The primers were designed to recognize both species-specific and common families. **b**, Positions of the amplicons in L1Hs consensus sequence and the number of *in silico* PCR hits on the human and chimp genomes. **c**, qRT–PCR analysis using primers for different regions of L1 element show higher levels of L1 RNA in NHP iPS cells regardless of the L1 region tested: 5′ UTR, ORF1 and ORF2 (mean ± s.e.m.;

$n = 3$ biological replicates; *$P < 0.01$ between indicated groups, *t*-test). **d**–**g**, Quantification of L1 elements in human and chimpanzee genomes using a population divergence model. Number of L1 elements found in the human and chimpanzee genomes for families: L1PA4 (**d**), L1PA3 (**e**), L1PA2 (**f**) and L1Pt and L1Hs (**g**) plotted as a histogram relative to their divergence (number of mutations relative to the canonical element). The standard deviation describes the differences in L1 density based on the sampling of different genomic regions and represents the variability of L1 coverage across the genomes (see Methods).
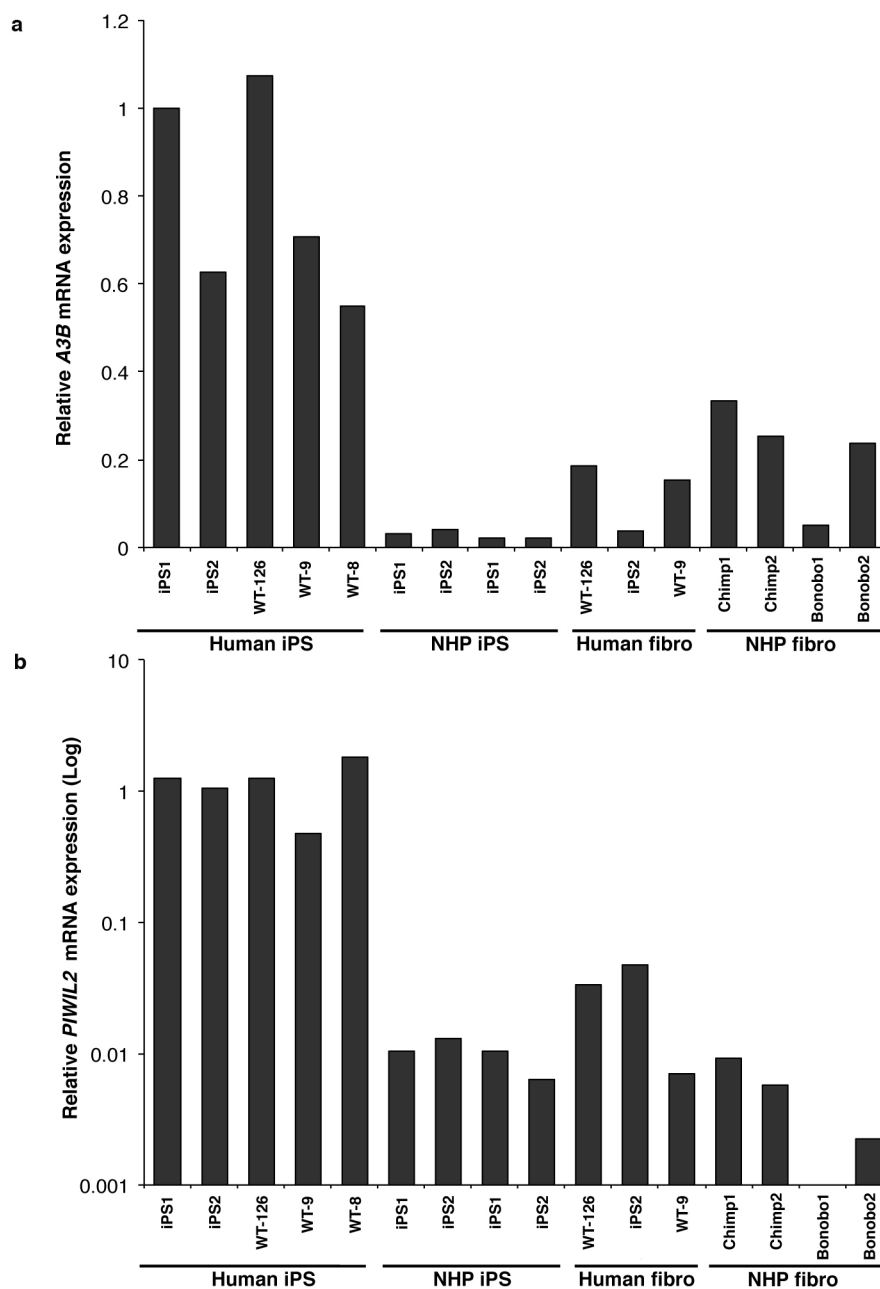
Extended Data Figure 9 | Relative *A3B* and *PIWIL2* mRNA levels in iPS cells and fibroblasts. Relative expression of *A3B* (**a**) and *PIWIL2* (**b**) in human and NHP iPS cell lines, and the available source fibroblasts from which iPS cells were derived. mRNA levels were normalized to *GAPDH* and shown relative to human iPS cell line 1.

Extended Data Table 1 | List of primers used in this study

Primers	Sequence	Use
Nanog-F	5'- CCTATGCCTGTGATTTGTGG -3'	PCR
Nanog-R	5'- CTGGGACCTTGTCTTCCTTT -3'	PCR
AFP-F	5'- AAAAGCCCACTCCAGCATC -3'	PCR
AFP-R	5'- CAGACAATCCAGCACATCTC -3'	PCR
Musashi-F	5'- AAAGGAGGTGATGTCGCCAA -3'	PCR
Musashi-R	5'- TGGTCCGTAGGCAGTGAGA -3'	PCR
Brachyury-F	5'- GCCCTCTCCCTCCCCTCCACGCACAG -3'	PCR
Brachyury-R	5'- CGGCGCCGTTGCTCACAGACCACAGG -3'	PCR
B-Actin-F	5'- TGTTTTCTGCGCAAGTTAGGTTTT -3'	PCR
B-Actin-R	5'- GCCGACAGGATGCAGAAGGAGAT -3'	PCR
APOBEC3B (20-40)	5'-GCGGGACAGGGACAAGCGTAT-3'	Cloning
APOPEC3B (1250-1228)	5'-CTGCTCAACCCAGGTCTCTGCCT-3'	Cloning
APOBEC3B (19-41)	5'-AGCGGGACAGGGACAAGC GTATC-3'	Cloning
APOBEC3B(1309-1288)	5'-AGCTGGAGATGGTGGTGAACGG-3'	Cloning
L1Pt ch7 11 F	5'-TTGCAGGTACTCTGAGCTTCAC-3'	Cloning
L1Pt ch7 11 R	5'-AAGGAGAAGCACCTGCATGA-3'	Cloning
Not-L1 F	5'-ATAAGAATGCGGCCGCGGGGGAGGAGCCAAGATG-3'	Cloning
XhoINotI L1 5UTR	5'-CCGCTCGAGCGGCCGCGGGGGAGGAG-3'	Cloning
L1 5UTRhindIIIATG	5"-TTTTTAAGCTTCCATCTTTGTGGTTTTATCTAC-3"	Cloning
APOBEC3B-F	5'-CGCCAGACCTACTTGTGCTAT-3'	qPCR
APOBEC3B-R	5'-CATTTGCAGCGCCTCCTTAT-3'	qPCR
GAPDH-F	5'- CATGTTCCAATATGATTCCACC-3'	qPCR
GAPDH-R	5'- CTCCACGACGTACTCAGCG-3'	qPCR
PIWIL2-F	5'- TTGTGGACAGCCTGAAGCTA -3'	qPCR
PIWIL2-R	5'- CCATCAGACACTCCATCACG -3'	qPCR
L1 5'UTR set1-R	5'-AAGATGGCCGAATAGGAACA-3'	qPCR
L1 5'UTR set1-R	5'-GATGAACCCGGTACCTCAGA-3'	qPCR
L1 5'UTR set2-R	5'-GAGATCTGAGAACGGGCAGA-3'	qPCR
L1 5'UTR set2-R	5'-AGCTGCAGGTCTGTTGGAAT-3'	qPCR
L1 ORF1 set1-F	5'-GCTACGGGAGGACATTCAAA-3'	qPCR
L1 ORF1 set1-R	5'-TTCAGCTCCATCAGCTCCTT-3'	qPCR
L1 ORF1 set2-F	5'-ATGAGCAAAGCCTCCAAGAA-3'	qPCR
L1ORF1 set2-R	5'-TTCTCCCCATCACTTTCAGG-3'	qPCR
L1 ORF2 set1-F	5'-TGACAAACCCACAGCCAATA-3'	qPCR
L1 ORF2 set1-R	5'-CCCTGTCTTGTGCCAGTTTT-3'	qPCR
L1 ORF2 set2-F	5'-TGGAGGCATCACACTACCTG-3'	qPCR
L1 ORF2 set2-R	5'-ATGCGGCATTATTTCTGAGG-3'	qPCR
Actin-F	5'- TACAATGAGCTGCGTGTGG-3'	qPCR
Actin-R	5'- TAGCACAGCCTGGATAGCAA-3'	qPCR
GFP F2	5'- GGGTGTTCTGCTGGTAGTGG-3'	qPCR
GFP R2	5'- TATATCATGGCCGACAAGCA-3'	qPCR
PURO F	5'- CTCGACATCGGCAAGGTGTG-3'	qPCR
PURO R	5'- GCCTTCCATCTGTTGCTGCG-3'	qPCR
APOBEC3A	TaqMan Assay (Life Technologies) Hs00377444	qPCR
APOBEC3B	TaqMan Assay (Life Technologies) Hs00358981	qPCR
APOBEC3C	TaqMan Assay (Life Technologies) Hs00828074	qPCR
APOBEC3D	TaqMan Assay (Life Technologies) Hs00537163	qPCR
APOBEC3F	TaqMan Assay (Life Technologies) Hs01665324	qPCR
APOBEC3G	TaqMan Assay (Life Technologies) Hs00222415	qPCR
APOBEC3H	TaqMan Assay (Life Technologies) Hs00962174	qPCR
PIWIL2	TaqMan Assay (Life Technologies) Hs01032720	qPCR
GAPDH	TaqMan Assay (Life Technologies) Hs03929097	qPCR
HPRT	TaqMan Assay (Life Technologies) Hs01003267	qPCR

LETTER

Cell intrinsic immunity spreads to bystander cells via the intercellular transfer of cGAMP

Andrea Ablasser[1], Jonathan L. Schmid–Burgk[1], Inga Hemmerling[1], Gabor L. Horvath[2], Tobias Schmidt[1], Eicke Latz[2,3] & Veit Hornung[1]

The innate immune defence of multicellular organisms against microbial pathogens requires cellular collaboration. Information exchange allowing immune cells to collaborate is generally attributed to soluble protein factors secreted by pathogen-sensing cells. Cytokines, such as type I interferons (IFNs), serve to alert non-infected cells to the possibility of pathogen challenge[1]. Moreover, in conjunction with chemokines they can instruct specialized immune cells to contain and eradicate microbial infection. Several receptors and signalling pathways exist that couple pathogen sensing to the induction of cytokines, whereas cytosolic recognition of nucleic acids seems to be exquisitely important for the activation of type I IFNs, master regulators of antiviral immunity[2]. Cytosolic DNA is sensed by the receptor cyclic GMP-AMP (cGA-MP) synthase (cGAS), which catalyses the synthesis of the second messenger cGAMP(2′-5′)[3-8]. This molecule in turn activates the endoplasmic reticulum (ER)-resident receptor STING[9-11], thereby inducing an antiviral state and the secretion of type I IFNs. Here we find in murine and human cells that cGAS-synthesized cGAMP (2′-5′) is transferred from producing cells to neighbouring cells through gap junctions, where it promotes STING activation and thus antiviral immunity independently of type I IFN signalling. In line with the limited cargo specificity of connexins, the proteins that assemble gap junction channels, most connexins tested were able to confer this bystander immunity, thus indicating a broad physiological relevance of this local immune collaboration. Collectively, these observations identify cGAS-triggered cGAMP(2′-5′) transfer as a novel host strategy that serves to rapidly convey antiviral immunity in a transcription-independent, horizontal manner.

On recognition of virus-derived nucleic acids, innate immune signalling initiates cell-autonomous antiviral effector mechanisms that aim to block viral propagation. Moreover, virus-infected cells alert non-infected neighbouring cells, a process largely attributed to the *de novo* expression and secretion of cytokines and chemokines. At the same time, a few reports have documented the phenomenon of cytokine-independent activation of bystander cells via gap junctions in the context of bacterial infection[12], irradiation[13] or DNA transfection[14]. However, the molecular mechanisms responsible for these effects remained elusive.

The finding that pattern sensing relies on a specific intermediate messenger molecule to activate a second receptor is unique in innate immunity, thus raising the question whether cGAMP(2′-5′)-mediated information transduction might provide organisms with an advantage over the use of a canonical, cell-autonomous signal transduction pathway[15].

Activation of STING triggers its oligomerization into a supramolecular complex and its translocation from the ER to a perinuclear compartment[16], a process that can be monitored at the single-cell level using fluorescence microscopy. To characterize the molecular mechanism of the cGAS–STING pathway better, we used HEK cells stably transduced with an amino-terminally mCherry-tagged STING construct (HEK STING)[17]. As expected, transient overexpression of cGAS–GFP in HEK STING cells led to phosphorylation of IRF3 and re-localization of STING to perinuclear complexes (Fig. 1a,
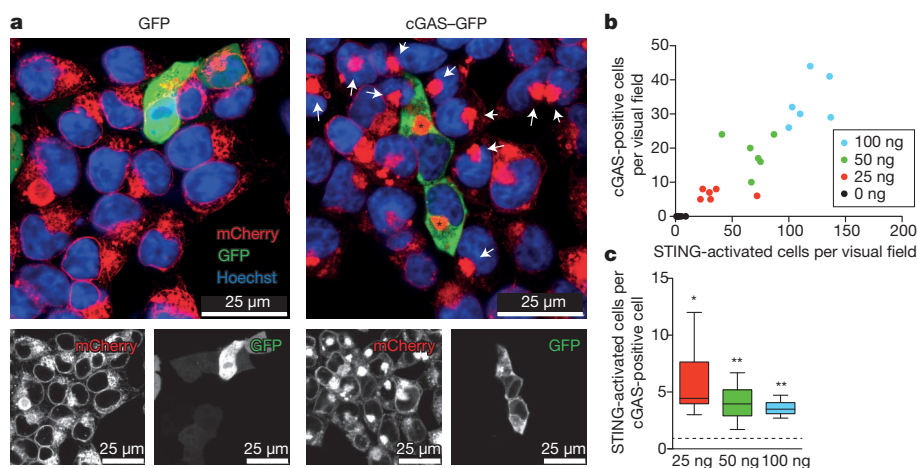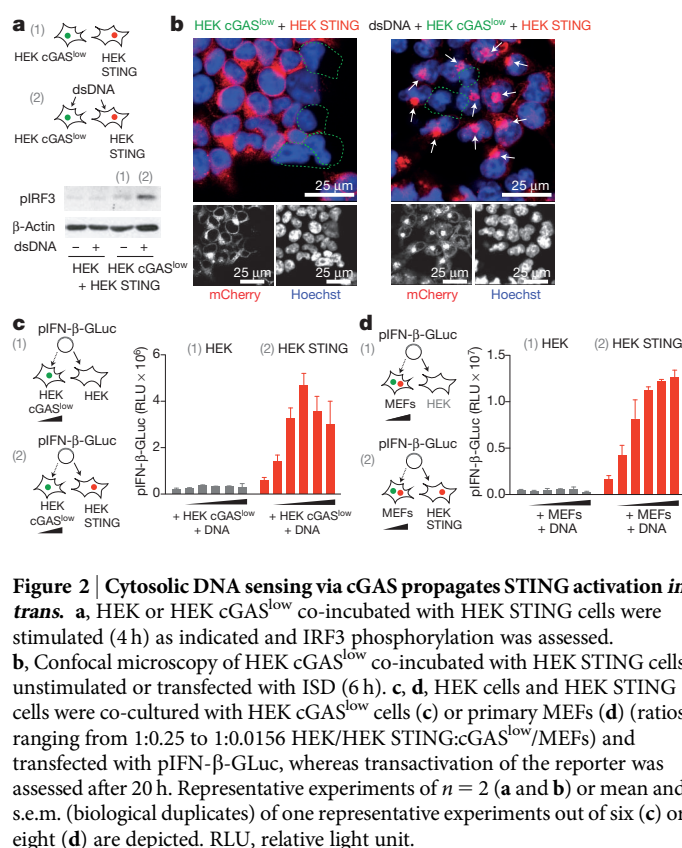

Figure 1 | cGAS overexpression activates STING in adjacent cells.
a, Confocal microscopy of HEK STING cells 20 h after transfection with GFP (left) or a cGAS–GFP (right). Asterisks and arrows highlight STING complexes in GFP-positive cells and bystander cells. **b, c**, HEK STING cells were transfected with varying amounts of cGAS–GFP as indicated. The number of GFP-positive cells is plotted against the number of activated HEK STING cells ($y = 0.27x$, $R^2 = 0.84$) (**b**) and the respective ratio of STING-activated cells over cGAS-expressing cells is depicted. Data are depicted as box plots with whiskers indicating minimum and maximum (**c**). One representative experiment out of two independent experiments is shown. *$P < 0.05$, **$P < 0.01$.

[1]Institute for Clinical Chemistry and Clinical Pharmacology, University Hospital, University of Bonn, 53127 Bonn, Germany. [2]Institute of Innate Immunity, University Hospital, University of Bonn, 53127 Bonn, Germany. [3]Department of Medicine, Division of Infectious Diseases and Immunology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

Figure 2 | Cytosolic DNA sensing via cGAS propagates STING activation *in trans*. a, HEK or HEK cGAS^{low} co-incubated with HEK STING cells were stimulated (4 h) as indicated and IRF3 phosphorylation was assessed. **b**, Confocal microscopy of HEK cGAS^{low} co-incubated with HEK STING cells unstimulated or transfected with ISD (6 h). **c, d**, HEK cells and HEK STING cells were co-cultured with HEK cGAS^{low} cells (**c**) or primary MEFs (**d**) (ratios ranging from 1:0.25 to 1:0.0156 HEK/HEK STING:cGAS^{low}/MEFs) and transfected with pIFN-β-GLuc, whereas transactivation of the reporter was assessed after 20 h. Representative experiments of n = 2 (**a** and **b**) or mean and s.e.m. (biological duplicates) of one representative experiments out of six (**c**) or eight (**d**) are depicted. RLU, relative light unit.

asterisks and data not shown). Surprisingly, we also observed STING translocation in cells that lacked cGAS–GFP expression, but that were located adjacent to cGAS-expressing cells (Fig. 1a, arrows). In contrast, the cell-permeable STING activator CMA induced homogenous STING clustering (see below)[17], indicating that stimulation of surrounding cells occurs via an event that is spatially and temporally linked to cGAS activity. Quantifying cGAS expression next to STING

activation revealed an approximately fourfold higher number of STING-activated cells compared to cGAS-expressing cells (Fig. 1b, c).

To assess the function of cGAS as a DNA receptor, we next generated monoclonal HEK cGAS cells with either high or low constitutive expression of cGAS. As expected, a cell clone with high cGAS expression (HEK cGAS*) induced spontaneous activation of STING and IRF3 phosphorylation in bystander cells (Extended Data Fig. 1 and data not shown). In contrast, a monoclonal cell line with low cGAS expression (HEK cGAS^{low}) additionally required DNA stimulation to exert STING and subsequent IRF3 activation in bystander cells (Fig. 2a, b). Moreover, titrating the number of HEK cGAS^{low} cells on top of STING competent cells in conjunction with DNA transfection revealed a dose-dependent increase in IFN-β promoter transactivation (Fig. 2c and Extended Data Fig. 2a). This bystander STING activation phenomenon was also observed when HEK STING cells were co-incubated with DNA-stimulated murine embryonic fibroblasts (MEFs) that are inherently competent for cGAS (Fig. 2d and Extended Data Fig. 2b). Of note, knockdown of cGAS in MEFs markedly decreased *in trans* activation of HEK STING cells following DNA stimulation (Extended Data Fig. 2c–f). Moreover, switching donor and recipient cells showed the same effect: HEK cGAS* but not unmodified HEK cells transactivated MEFs and the murine cell line LL171 in a STING-dependent fashion, indicating that cGAS-dependent STING activation *in trans* was conserved across species (see below). Notably, this phenomenon of bystander cell activation was not observed when expressing an RNA-polymerase-III-driven RIG-I stimulatory RNA molecule[18]: whereas cell-intrinsic RIG-I activation was observed under these conditions, no bystander activation could be detected (Extended Data Fig. 3).

Separating donor and recipient cells via a trans-well system completely blunted bystander cell activation (Extended Data Fig. 4), indicating that a cell-to-cell contact-dependent transfer mechanism was responsible for conveying the IRF3 activating signal[14]. When we loaded HEK cGAS* cells with the low-molecular-weight dye calcein as a tracer, we observed transfer of calcein from HEK cGAS* cells into HEK STING cells that were in direct or indirect contact (Fig. 3a). Most notably, calcein transfer coincided with STING activation in the recipient cells, indicating a physical connection of signal transduction (Fig. 3a and
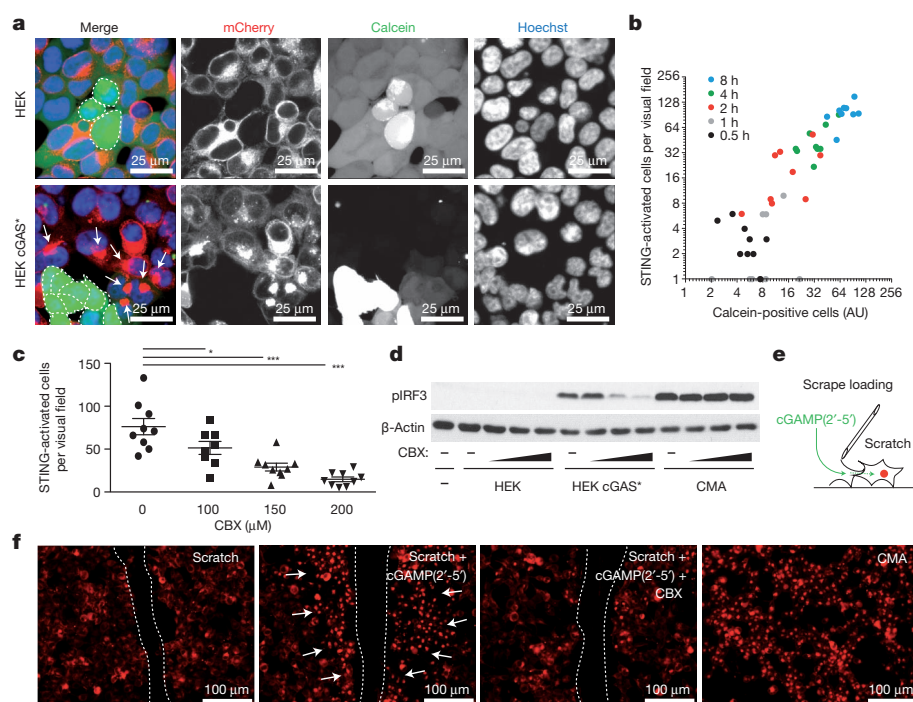
Figure 3 | cGAS-produced cGAMP(2′-5′) passes through gap junctions to trigger STING activation in bystander cells. a, Confocal microscopy of HEK cells and HEK cGAS* cells loaded with calcein and added to HEK STING cells for 4 h. **b**, Co-culturing was performed as in **a** and after 0–8 h HEK STING cells were analysed by fluorescence microscopy for STING aggregation in nine independent visual fields. A dot-blot diagram correlating calcein-positive HEK STING cells with STING aggregate formation is presented ($y = 1.27x$, $R^2 = 0.85$). AU, arbitrary units. **c**, HEK STING cells were co-cultured with HEK cGAS* in the presence of CBX as indicated for 4 h and studied for STING activation in nine independent visual fields. Data are presented as mean and s.e.m. **d**, IRF3 phosphorylation in HEK STING cells with HEK cells, with HEK cGAS* cells or with CMA in the presence or absence of CBX (100 μM, 150 μM and 200 μM) (3 h). **e**, The scrape loading technique. **f**, Fluorescence images of wounded HEK STING cells incubated with nothing, cGAMP(2′-5′) or cGAMP(2′-5′) with 150 μM CBX. HEK STING cells with CMA served as control (dashed line, scratch margins; arrows, STING complexes). One representative experiment out of two independent experiments is shown (**a–d, f**).
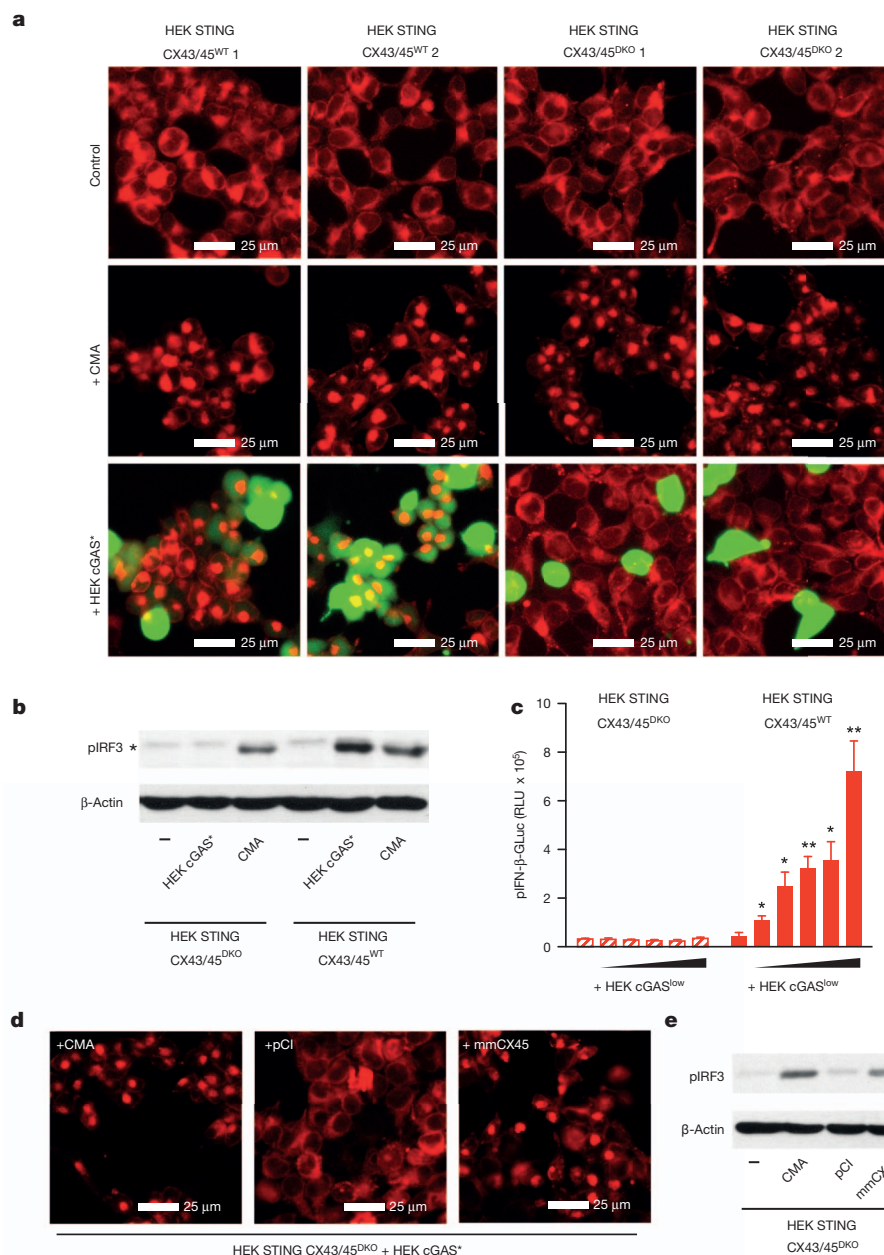*$P < 0.05$, ***$P < 0.001$.

Figure 4 | Connexin 43 and 45 mediate cGAMP(2′-5′) transfer in HEK STING cells. **a**, Fluorescence microscopy of HEK STING CX43/45[WT] and HEK STING CX43/45[DKO] cells left untreated, stimulated with CMA or co-cultured with calcein-loaded HEK cGAS* cells (ratio HEK/HEK STING:HEK cGAS* = 1:0.25) after 8 h. **b**, Phosphorylation of IRF3 in HEK STING CX43/45[WT] and HEK STING CX43/45[DKO] cells left untreated, co-incubated with HEK cGAS* cells or stimulated with CMA for 4 h (asterisk indicates nonspecific band). **c**, HEK STING CX43/45[WT] and HEK STING CX43/45[DKO] cells were co-cultured with HEK cGAS[low] cells (ratios from 1:0.5 to 1:0.0312 HEK STING CX43/45[WT] or HEK STING CX43/45[DKO] cells:cGAS[low]) transfected with pIFN-β-GLuc and luciferase activity was assessed after 20 h. Mean and s.e.m. (biological duplicates) of one representative experiment out of three independent experiments is shown. **d**, Fluorescence microscopy of HEK STING CX43/45[DKO] cells co-cultured with HEK cGAS* cells and stimulated with CMA, transfected with empty vector (pCI) or an expression vector for murine CX45 (mmCX45, 20 h). **e**, Co-cultures from **d** were analysed for phosphorylation of IRF3. One representative experiment out of two independent experiments is shown (**a**, **b**, **d**, **e**). *$P < 0.05$, **$P < 0.01$.

Supplementary Videos 1 and 2). Indeed, quantitative analysis of calcein transfer and STING activation over time revealed a strong correlation of the two processes (Fig. 3b). These results indicated the involvement of gap junctions, which represent a well-established route of cell-contact-dependent intercellular communication. We therefore tested the impact of carbenoxolone (CBX), a well-characterized inhibitor of connexin function and thus gap junctions. CBX treatment potently inhibited calcein transfer from HEK cGAS* cells to HEK STING cells in a dose-dependent fashion and also blocked STING activation and IRF3 phosphorylation (Fig. 3c, d and Extended Data Fig. 5). Gap junctions allow the passage of small molecules below 1 kDa

between cells, whereas larger biomolecules or proteins are precluded from this means of intercellular connection[19]. The fact that cGAS expression per se led to bystander STING activation indicated that cGAMP(2′-5′) was the second messenger molecule transported across gap junctions. To test this hypothesis we made use of scrape loading[20], which is a well-established technique to study gap junction intercellular communication *in vitro*. In this assay a cellular monolayer is wounded by a cut allowing the extracellular space to gain access to the cytoplasm of lacerated cells, which are still coupled to neighbouring cells through gap junctions (Fig. 3e). Supplying cGAMP(2′-5′) to scratched HEK STING cells led to rapid and strong STING activation
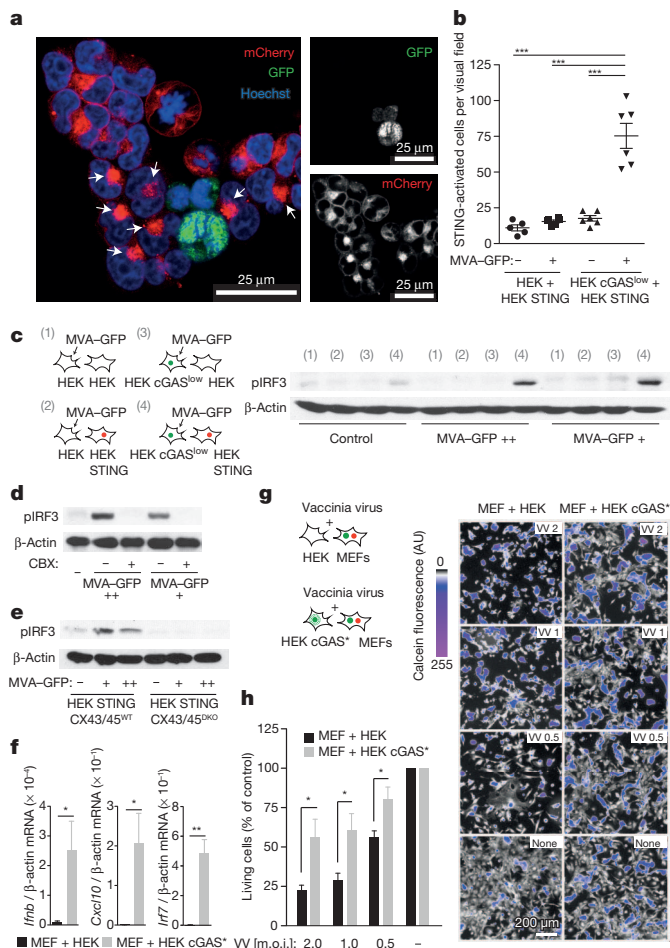
Figure 5 | Vaccina virus triggers STING-dependent antiviral immunity in bystander cells. **a**, **b**, HEK cGAS[low] cells or HEK cells were infected with MVA–GFP, washed and then loaded onto HEK STING cells for 8 h. **a**, Confocal microscopy of HEK STING cells co-incubated with MVA-infected HEK cGAS cells (arrows, STING activation in bystander cells). **b**, Quantification of STING activation in response to MVA-infected HEK cGAS[low] cells or MVA-infected HEK cells. **c**, HEK cells or HEK cGAS[low] cells were MVA-infected or left untreated, added onto HEK cells or HEK STING cells and studied for IRF3 phosphorylation (viral particles per ml: $++ = 3.2 \times 10^7$; $+ = 1.6 \times 10^7$). **d**, **e**, Experiments as in **a** in the presence or absence of CBX 150 μM (**d**) or using HEK STING CX43/45[WT] and HEK STING CX43/45[DKO] cells as responder cells (**e**). **f**, HEK cGAS* cells were co-incubated with MEFs and after 14 h mouse *Ifnb* mRNA, mouse *Cxcl10* mRNA and mouse *Irf7* mRNA were assessed by qPCR. **g**, **h**, HEK cGAS* cells were co-incubated with MEFs and after 12 h vaccinia virus was added (multiplicity of infection (m.o.i.): 2–0.5). Twenty-four hours later cell survival was analysed. Visual fields of one representative experiment are depicted (**g**, right panel) and mean and s.e.m. of three independent experiments are summarized (**h**). VV, vaccinia virus. One representative experiment out of two (**a**, **b**, **e**) or three (**c**, **d**) independent experiments are shown or mean and s.e.m. of $n = 5$ independent experiments is presented (**f**). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

along the margins of the laceration and CBX treatment abrogated this effect (Fig. 3f and Extended Data Fig. 6).

Gap junctions are formed by connexin proteins that assemble into clusters of hundreds of intercellular channels, thus physically connecting the cytoplasm of neighbouring cells. The connexin family of proteins consists of 20 members in the mouse and 21 members in the human system, with overlapping yet distinctive cellular distribution patterns[19]. HEK cells have been reported to express connexins 43 (CX43) and 45 (CX45)[21,22]. Consequently, to investigate the mechanism of gap-junction-mediated transfer of cGAMP(2′-5′), we simultaneously targeted CX43 and CX45 in HEK STING cells using the CRISPR/Cas9 system[23,24]. We thus generated two CX43/CX45-competent (wild

type (WT)) and two CX43/CX45-deficient (double knockout (DKO)) HEK STING cell lines (HEK STING CX43/45[WT] cells and HEK STING CX43/45[DKO] cells) (Extended Data Fig. 7). HEK STING CX43/45[DKO] cells showed normal responsiveness towards CMA, indicating that STING responses were intact in these cells (Fig. 4a, b). However, co-culture of HEK STING CX43/45[DKO] cells with calcein-loaded HEK cGAS* cells did not result in dye transfer and, at the same time, STING activation, IRF3 phosphorylation and transactivation of an IFN-β reporter were completely abrogated (Fig. 4a–c). In line with this, DNA-stimulated MEFs only transactivated gap-junction-competent HEK STING CX43/45[WT] cells but not HEK STING CX43/45[DKO] cells (data not shown). Moreover, scrape loading of HEK STING CX43/45[DKO] cells with cGAMP (2′-5′) did not give rise to extended patches of STING-activated cells (Extended Data Fig. 8a). Reconstitution of CX45 and CX43 in HEK STING CX43/45[DKO] cells restored bystander STING activation and the resulting phosphorylation of IRF3 (Fig. 4d, e and Extended Data Fig. 8b; data not shown) and, at the same time, overexpression of five out of six additional distinct human connexin family members restored bystander cell activation in HEK STING CX43/45[DKO] cells as well (Extended Data Fig. 8b). This mutual potential for complementation is in line with the notion that connexins have only limited cargo specificity and thus can functionally compensate for other family members[19].

We next sought to address the physiological relevance of this phenomenon in the context of infection with a DNA virus known to activate STING. With this aim, we infected HEK cGAS[low] cells with a replication-incompetent vaccinia virus strain encoding nucleus-targeted GFP (MVA–GFP, modified vaccinia Ankara). Three hours after infection, HEK cGAS[low] cells were washed and co-cultured with HEK STING cells. HEK STING cells adjacent to MVA-infected cells showed prominent STING clustering (Fig. 5a, b) as well as robust IRF3 phosphorylation (Fig. 5c). Of note, this MVA-initiated antiviral signalling was dependent on both cGAS expression in donor cells as well as STING expression in recipient cells. Consistent with our earlier results, MVA-initiated bystander antiviral response was blocked by CBX (Fig. 5d) and was absent in HEK STING cells deficient for CX43/45 (Fig. 5e). Similarly, co-culturing MVA-infected MEFs induced upregulation of IFN-β in HEK STING cells, but not in unmodified HEK cells, and required functional gap junctions (Extended Data Fig. 9). Finally, we tested whether this bystander STING activation conferred antiviral protection. To this end, we made use of a replication-competent strain of vaccinia virus (Western Reserve) that induces rapid cell death in primary MEFs. Co-incubating MEFs with HEK cGAS* cells induced strong expression of antiviral genes in MEFs (Fig. 5f) and led to a marked increase in viral resistance of MEFs infected with vaccinia virus, as observed by an increase in cell viability (Fig. 5g, h).

We provide a clear delineation of a unique *in trans* innate immune signalling mechanism that comprises cGAMP(2′-5′) being produced by cGAS in the sensing cell, which is delivered through gap junctions to bystander cells, leading to remote STING activation and subsequent antiviral immunity (Extended Data Fig. 10). Compared to the transcriptionally regulated paracrine activation of bystander cells, for example, via type I IFNs, the gap-junction-dependent transfer of cGAMP(2′-5′) might provide several key advantages to the host. Foremost, type I IFN-dependent induction of antiviral immunity in bystander cells takes considerably longer, given the requirement of a *de novo* transcription and translation event within the sensing cell. In addition, many viruses block type I IFN induction within the infected cell at multiple interdependent levels, whereas cGAS-dependent cGAMP(2′-5′) synthesis only requires host ATP and GTP, thus exhibiting only a minimal target for virus-encoded inhibitory mechanisms. As such, a virus-infected, potentially compromised cell can still propagate and even amplify antiviral immunity relying on bystander cells connected through gap junctions.

Although bystander activation and signal amplification might prove beneficial for the restriction of viral infection, it might at the same time

aggravate disease manifestations in STING-dependent autoimmune syndromes (for example, Aicardi–Goutières syndrome)[25] and as such potentially provide a novel target for therapeutic strategies.

METHODS SUMMARY

Cell stimulation and transient transfection. Co-cultures of HEK cells and HEK STING cells with MEFs and HEK cGAS*/cGAS^{low} cells were transfected with Lipofectamine and 200 ng of reporter plasmid, which served as stimulus and reporter at the same time. Unless otherwise indicated transient overexpression of cDNA constructs was performed using GeneJuice (Novagen) according to the manufacturer's instructions. In some titration experiments pCI empty vector was used as stuffer plasmid.

Epifluorescence and confocal fluorescence microscopy. For epifluorescence microscopy HEK STING cells were seeded at a density of 2.5×10^4 cells per well in poly-L-lysine-coated 96-well plates. HEK cGAS* cells (8,000 per well) pre-incubated with $2\,\mu g\,ml^{-1}$ calcein-AM for 20–60 min at 37 °C were added on top of HEK STING cells. Images were collected using a Zeiss Observer.Z1 inverted microscope with ×20 long-distance objective 8 h after stimulation or co-culture if not otherwise indicated. Confocal microscopy was performed with living cells on a Leica SP5 SMD confocal microscope with a ×63 water-immersion objective at 37 °C. For nucleus staining, cells were incubated with Hoechst 34580 dye (10 μg ml^{-1}) 30 min before imaging.

Scrape loading. HEK STING cells were seeded at a density of 2.5×10^5 cells ml^{-1} in 96-well plates. Monolayers of cells were manually wounded by six scratches per well using an 18G needle.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Sadler, A. J. & Williams, B. R. Interferon-inducible antiviral effectors. *Nature Rev. Immunol.* **8,** 559–568 (2008).
2. Goubau, D., Deddouche, S. & Reis, E. S. C. Cytosolic sensing of viruses. *Immunity* **38,** 855–869 (2013).
3. Wu, J. *et al.* Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science* **339,** 826–830 (2013).
4. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339,** 786–791 (2013).
5. Gao, P. *et al.* Cyclic [G(2′,5′)pA(3′,5′)p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell* **153,** 1094–1107 (2013).
6. Diner, E. J. *et al.* The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.* **3,** 1355–1361 (2013).
7. Ablasser, A. *et al.* cGAS produces a 2′-5′-linked cyclic dinucleotide second messenger that activates STING. *Nature* **498,** 380–384 (2013).
8. Zhang, X. *et al.* Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell* **51,** 226–235 (2013).
9. Ishikawa, H. & Barber, G. N. STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature* **455,** 674–678 (2008).
10. Zhong, B. *et al.* The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. *Immunity* **29,** 538–550 (2008).
11. Sun, W. *et al.* ERIS, an endoplasmic reticulum IFN stimulator, activates innate immune signaling through dimerization. *Proc. Natl Acad. Sci. USA* **106,** 8653–8658 (2009).
12. Kasper, C. A. *et al.* Cell-cell propagation of NF-κB transcription factor and MAP kinase activation amplifies innate immunity against bacterial infection. *Immunity* **33,** 804–816 (2010).
13. Hamada, N., Matsumoto, H., Hara, T. & Kobayashi, Y. Intercellular and intracellular signaling pathways mediating ionizing radiation-induced bystander effects. *J. Radiat. Res.* **48,** 87–95 (2007).
14. Patel, S. J., King, K. R., Casali, M. & Yarmush, M. L. DNA-triggered innate immune responses are propagated by gap junction communication. *Proc. Natl Acad. Sci. USA* **106,** 12867–12872 (2009).
15. Ablasser, A. & Hornung, V. DNA sensing unchained. *Cell Res.* **23,** 585–587 (2013).
16. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461,** 788–792 (2009).
17. Cavlar, T., Deimling, T., Ablasser, A., Hopfner, K. P. & Hornung, V. Species-specific detection of the antiviral small-molecule compound CMA by STING. *EMBO J.* **32,** 1440–1450 (2013).
18. Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nature Immunol.* **10,** 1065–1072 (2009).
19. Laird, D. W. Life cycle of connexins in health and disease. *Biochem. J.* **394,** 527–543 (2006).
20. Juul, M. H., Rivedal, E., Stokke, T. & Sanner, T. Quantitative determination of gap junction intercellular communication using flow cytometric measurement of fluorescent dye transfer. *Cell Adhes. Commun.* **7,** 501–512 (2000).
21. Butterweck, A., Gergs, U., Elfgang, C., Willecke, K. & Traub, O. Immunochemical characterization of the gap junction protein connexin45 in mouse kidney and transfected human HeLa cells. *J. Membr. Biol.* **141,** 247–256 (1994).
22. Langlois, S., Cowan, K. N., Shao, Q., Cowan, B. J. & Laird, D. W. Caveolin-1 and -2 interact with connexin43 and regulate gap junctional intercellular communication in keratinocytes. *Mol. Biol. Cell* **19,** 912–928 (2008).
23. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).
24. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
25. Gall, A. *et al.* Autoimmunity initiates in nonhematopoietic cells and progresses via lymphocytes in an interferon-dependent autoimmune disease. *Immunity* **36,** 120–131 (2012).

Author Contributions A.A., J.L.S.-B., I.H. and V.H. designed experiments and analysed the data. A.A., I.H., J.L.S.-B., G.L.H. and E.L. performed experiments. J.L.S.-B. and T.S. developed the CRISPR/Cas9 targeting strategy. A.A., J.L.S.-B. and V.H. wrote the manuscript. V.H. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.H. (veit.hornung@uni-bonn.de).

METHODS

Reagents and plasmids. DNA oligonucleotides corresponding to the 45-base-pair (bp) long dsDNA interferon stimulatory DNA (ISD) (sense sequence 5′-TACAGATCTACTAGTGATCTATGACTGATCTGTACATGATCTACA-3′)[26] were obtained from Metabion and annealed in PBS. 10-carboxymethyl-9-acridanone, dynasore, poly-L-lysine and carbenoxolone were from Sigma-Aldrich. Calcein-AM and Hoechst were from Invitrogen. Recombinant murine IFN-α was from PBL interferon source. Cyclic di-GMP was obtained from Biolog. cGAMP(2′-5′) was generated via in vitro enzymatic assay and purified as described previously[7]. Expression plasmids encoding for human GFP-tagged cGAS[27] and for an IFN-inducing stimulatory shRNA molecule[18] were previously described. Expression plasmid encoding for GFP is based on pEFBOS. Plasmids encoding human connexins were obtained from Thermo Scientific (Precision LentiORFs). Murine connexin expression constructs were provided by K. Willecke.

Cell culture. HEK cells (HEK 293T cells throughout the study), primary mouse embryonic fibroblasts (MEFs) and LL171 cells[28] (L929 cells containing a stable IFN-stimulated response element-luciferase reporter plasmid (ISRE-Luc)) were cultured in DMEM supplemented with 10% (v/v) FCS, sodium pyruvate (all Life Technologies) and ciprofloxacin (Bayer Schering Pharma). The HEK cell line stably expressing murine STING, which contains an N-terminal mCherry-tag, was previously described[17], whereas HEK cells expressing murine cGAS were generated by retroviral transduction using the pRP system. No mycoplasma contamination was detected in regular screenings of our cell lines.

Cell stimulation and transient transfection. Direct stimulation of HEK cells, HEK STING cells or HEK cGAS (3.5×10^5 cells ml^{-1} in 96-well plates) was performed by transfecting cyclic di-GMP ($2 \mu g$ ml^{-1}) or dsDNA ($1.33 \mu g$ ml^{-1}) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. CMA was added to the cells at a final concentration of $500 \mu g$ ml^{-1}. LL171 cells (0.15×10^6 per ml) were stimulated with recombinant IFN-α at a final concentration of 250 U ml^{-1}. DNA-stimulated co-cultures of HEK cells and HEK STING cells with MEFs and HEK cGAS*/cGAS^{low} cells were performed in a 96-well format and transfected with Lipofectamine and 200 ng of reporter plasmid, which served as stimulus and reporter at the same time. Unless otherwise indicated, transient overexpression of cDNA constructs was performed using GeneJuice (Novagen) according to the manufacturer's instructions. In some titration experiments pCI empty vector was used as stuffer plasmid. Unless otherwise indicated, cells were analysed 3–4 h after direct stimulation or 14–20 h after transfection of expression plasmids.

Immunoblotting. Cells were lysed in 1× Laemmli buffer and denatured at 95 °C for 5 min. Cell lysates were separated by 10% or 8% (for CX43) SDS–PAGE and transferred onto nitrocellulose membranes. Blots were incubated with anti-phospho-IRF3 (Cell Signaling Technology), anti-cGAS (Sigma), anti-CX45 (G-7) (Santa Cruz Biotechnology) or anti-CX43 (gift from K. Willecke) as primary and anti-rabbit-IgG-HRP and anti-mouse-IgG-HRP as secondary antibody or β-actin-IgG-HRP (all Santa Cruz Biotechnology).

Luciferase reporter assays. LL171 cells were lysed in 5× passive lysis buffer (Promega) for 10 min at room temperature. The total cell lysate was incubated with firefly luciferase substrate at a 1:1 ratio. Transactivation of a transiently expressed IFN-β Gaussia luciferase construct (pIFN-β-GLuc) was assessed in the supernatants 14–20 h after transfection with coelenterazine (2.2 μM) as substrate. Luminescence was measured on an EnVision 2104 Multilabel Reader (Perkin Elmer).

qPCR. RNA from cells was reverse transcribed using the RevertAid First Strand cDNA Synthesis kit (Fermentas) and quantitative PCR analysis was performed on an ABI 7900HT. All gene expression data are presented as relative expression to murine β-actin (murine cells) or human GAPDH (human cells). For human transcripts, the sequences were as follows: GAPDH forward 5′-GAGTCAACGGATTTGGTCGT-3′, GAPDH reverse 5′-GACAAGCTTCCCGTTCTCAG-3′; IFNB forward 5′-CAGCATCTGCTGGTTGAAGA-3′, IFNB reverse 5′-CATTACCTGAAGGCCAAGGA-3′; CXCL10 forward 5′-TCTGAATCCAGAATCGAAGG-3′, CXCL10 reverse 5′-CTCTGTGTGGTCCATCCTTG-3′. Primer sequences for murine IFN-β were as previously described[18]. For additional murine transcripts, the sequences were as follows: cGAS (Mb21d1) forward 5′-ACCGGACAAGCTAAAGAAGGTGCT-3′, cGAS reverse 5′-GCAGCAGGCGTTCCACAACTTTAT-3′; Sting (Tmem173) forward 5′-CACCTCTCTGAGCCTCAACC-3′, Sting reverse 5′-CCATCCACACAGGTCAACAG-3′; Irf7 forward 5′-GAAGACCCTGATCCTGGTGA-3′, Irf7 reverse 5′-CCAGGTCCATGAGGAAGTGT-3′; Cxcl10 forward 5′-AAGTGCTGCCGTCATTTTCT-3′, Cxcl10 reverse 5′-GTGGCAATGATCTCAACACG-3′; β-actin (Actb) forward 5′-AGCCATGTACGTAGCCATCC-3′, β-actin reverse 5′-CTCTCAGCTGTGGTGGTGAA-3′.

Epifluorescence and confocal fluorescence microscopy. For epifluorescence microscopy HEK STING cells were seeded at a density of 2.5×10^4 cells per well in poly-L-lysine-coated 96-well plates. HEK cGAS* cells (8,000 per well) pre-incubated with $2 \mu g$ ml^{-1} calcein-AM for 20–60 min at 37 °C were added on top of HEK STING cells. Calcein was chosen as the tracer of choice because of its comparable physiochemical properties to cGAMP(2′-5′). Images were collected using a Zeiss Observer.Z1 inverted microscope with ×1 (assessment of cell viability after vaccinia virus infection) or ×20 (STING activity assay in HEK cells) long-distance objective 8 h after stimulation or co-culture if not otherwise indicated. Image evaluation was performed as follows: STING aggregates were counted using the ImageJ plugin 'Cell Counter', whereas calcein dye transfer was quantified by measuring the area of the images with a green fluorescence value in between two arbitrary thresholds. The thresholds were chosen such as to exclude non-receiving acceptor cells as well as the very bright donor cells. Therefore, the resulting area measurements correspond only to acceptor cells having received calcein from neighbouring cells. Confocal microscopy was performed with living cells on a Leica SP5 SMD confocal microscope with a ×63 water-immersion objective at 37 °C. For nucleus staining, cells were incubated with Hoechst 34580 dye ($10 \mu g$ ml^{-1}) 30 min before imaging.

Scrape loading. HEK STING cells were seeded at a density of 2.5×10^5 cells ml^{-1} in 96-well plates. After 16 h cGAMP(2′-5′) was added to the medium to a final concentration of $50 \mu g$ ml^{-1}. Monolayers of cells were manually wounded by six scratches per well using an 18G needle. Images were acquired after 4–8 h.

MVA and vaccinia virus infection assay. HEK cells, HEK cGAS^{low} cells (both 3.2×10^5 cells per 12 wells) or MEFs (1.6×10^5 cells per 12 wells) were infected—if not otherwise indicated—with 3.2×10^7 or 1.6×10^7 viral particles per ml (for HEK cells) and 3.2×10^6 or 1.6×10^6 virus particles per ml (for MEFs) of MVA NP-S-GFP (MVA-GFP), which targets GFP to the nucleus of infected cells[29]. By applying these concentrations of viral particles, a homogenous expression of GFP was observed in more than 80% of recipient cells at the highest concentration. Three hours after infection, cells were washed three times and added onto HEK cells, HEK STING cells, HEK STING CX43/CX45^{WT} and HEK STING CX43/CX45^{DKO} cell lines, respectively (3.5×10^4 cells per 96-well or 4×10^5 cells per 12-well). In some experiments HEK STING cells were pre-treated with $150 \mu M$ CBX before co-culturing. After 6–8 h HEK cell co-cultures were either visualized by epifluorescence and confocal fluorescence microscopy or lysed for assessment of IRF3 phosphorylation by immunoblot. Co-cultures of infected MEFs and HEK cells or HEK STING cells were incubated overnight and induction of human IFN-β was measured via qPCR. Before infection with vaccinia virus (m.o.i. 2, 1 and 0.5) MEFs (1.5×10^4 cells per 96-well) were co-cultured with HEK cells or HEK cGAS* cells (5,000 per well) for 12 h. Cell survival in infected co-cultures was determined 24 h later via an epifluorescence-microscopy-based viability assay using calcein as a marker for viable cells. For these studies, calcein was added 1 h before microscopic analysis. Evaluation of images was performed using ImageJ by determining the surface area of monolayers of viable cells. In parallel induction of Ifnb, Cxcl10 and Irf7 in MEF–HEK cell or MEF–HEK STING cell co-cultures was quantified via qPCR.

siRNA experiments. siRNAs (Mission siRNA) against murine STING, described before[4], murine cGAS and control siRNA were purchased from Sigma and transfected into MEFs and LL171 using Lipofectamine 2000 (Invitrogen) at a final concentration of 50 nM: mmSTING 5′-CGAAAUAACUGCCGCCUCAdTdT-3′; mmcGAS#1 5′-GAUUGAGCUACAAGAAUAUdTdT-3′, mmcGAS#2 5′-GAGGAAAUCCGCUGAGUCAdTdT-3′; MissionsiRNA Universal Negative Control 1. Forty-eight hours after transfection cells were used for further experiments and knockdown of the indicated genes was verified by qPCR.

CRISPR/Cas9-mediated knockout cell-line generation. HEK STING cells were transfected in duplicates with 150 ng of a Cas9 expression plasmid together with 25 ng of two U6-gRNA expression plasmids specific for early coding exons of the human GJA1 and GJC1 genes. After 2 days, genome editing at both loci was verified by a T7EI endonuclease assay as described[30]. Limiting dilution cloning was performed by plating on average 0.8 cells in each well of six 96-well plates. After 10 days, growing clones were selected by bright-field microscopy and split to obtain two clonal duplicates. In one of the duplicate plates, 8,000 HEK cGAS* per well were co-seeded and cells were analysed with an epifluorescence microscope after 8 h to assess STING activation. Two clones not responsive as well as two responsive control clones were selected and the corresponding replicate cells were expanded for further experiments.

Deep-sequencing-based genotyping of connexin-deficient cells. Genomic DNA was isolated using a direct lysis buffer (0.2 mg ml^{-1} proteinase K, 1 mM CaCl$_2$, 3 mM MgCl$_2$, 1 mM EDTA, 1% Triton X-100, 10 mM Tris pH 7.5). The loci of interest were PCR-amplified and subsequently, using a secondary PCR, Illumina-compatible linkers and barcode sequences were added to the amplicons. The products of individual clones were pooled, gel- and silica-column purified, precipitated, quantified using a NanoDrop photospectrometer and used for an Illumina MiSeq 250 bases single read run using the v2 chemistry. FastQ data were analysed to call the allelic genotypes of the clones analysed.

Statistical analysis. On the basis of previous experience and expectations of biological effects, experiments that were assessed for statistical significance were typically performed 3–6 times (luciferase assays, qPCR studies) or 5–15 times (visual fields for microscopy studies). If not stated otherwise, data are presented as arithmetic means + s.e.m. and statistical analyses of the generally normally distributed data (Shapiro–Wilk normality test) were based on paired or unpaired *t*-tests, as appropriate. If unequal variances were observed for unpaired sample sets (*F* test for unequal variance), an unpaired *t*-test with Welch's correction was performed. Statistical analyses of normalized data were performed using a one-sample *t*-test. All data calculations were performed using GraphPad Prism.

26. Stetson, D. B. & Medzhitov, R. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* **24,** 93–103 (2006).
27. Civril, F. *et al.* Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498,** 332–337 (2013).
28. Uzé, G. *et al.* Domains of interaction between alpha interferon and its receptor components. *J. Mol. Biol.* **243,** 245–257 (1994).
29. Kastenmüller, W. *et al.* Peripheral prepositioning and local CXCL9 chemokine-mediated guidance orchestrate rapid memory CD8[+] T cell responses in the lymph node. *Immunity* **38,** 502–513 (2013).
30. Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Honing, K. & Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nature Biotechnol.* **31,** 76–81 (2013).
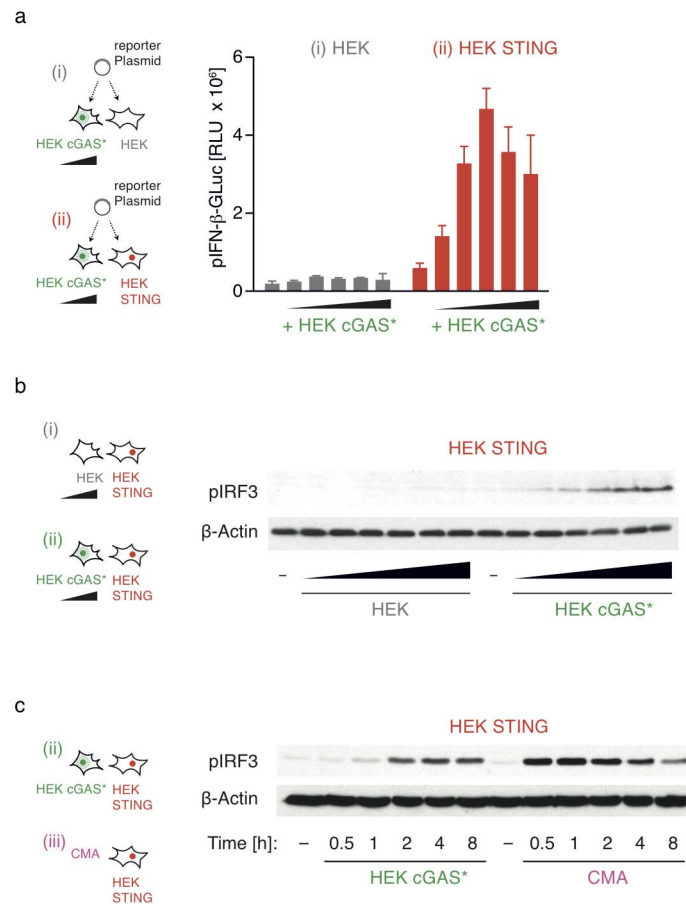
Extended Data Figure 1 | Stable overexpression of cGAS in HEK cells induces activation of HEK STING cells *in trans*. a, HEK cells and HEK STING cells were co-cultured with increasing amounts of HEK cGAS* cells (ratios ranging from 1:0.25 to 1:0.0156 HEK/HEK STING:HEK cGAS*). Co-cultures were transfected with pIFN-β-GLuc and after 20 h transactivation of the reporter construct was assessed. Mean and s.e.m. (biological duplicates) of one representative experiment out of two independent experiments are depicted. **b**, HEK STING cells were co-cultured with HEK cells or HEK cGAS* cells (ratios HEK STING:HEK/HEK cGAS* = 1:0.25) for 4 h and phosphorylation of IRF3 was determined in the cellular lysates by immunoblotting. **c**, Kinetics of IRF3 phosphorylation of HEK STING and HEK cGAS* co-cultures (ratio HEK STING:HEK cGAS* = 1:0.25) are depicted. CMA served as a control stimulus. Representative experiments of two independent experiments are shown (**b**, **c**).
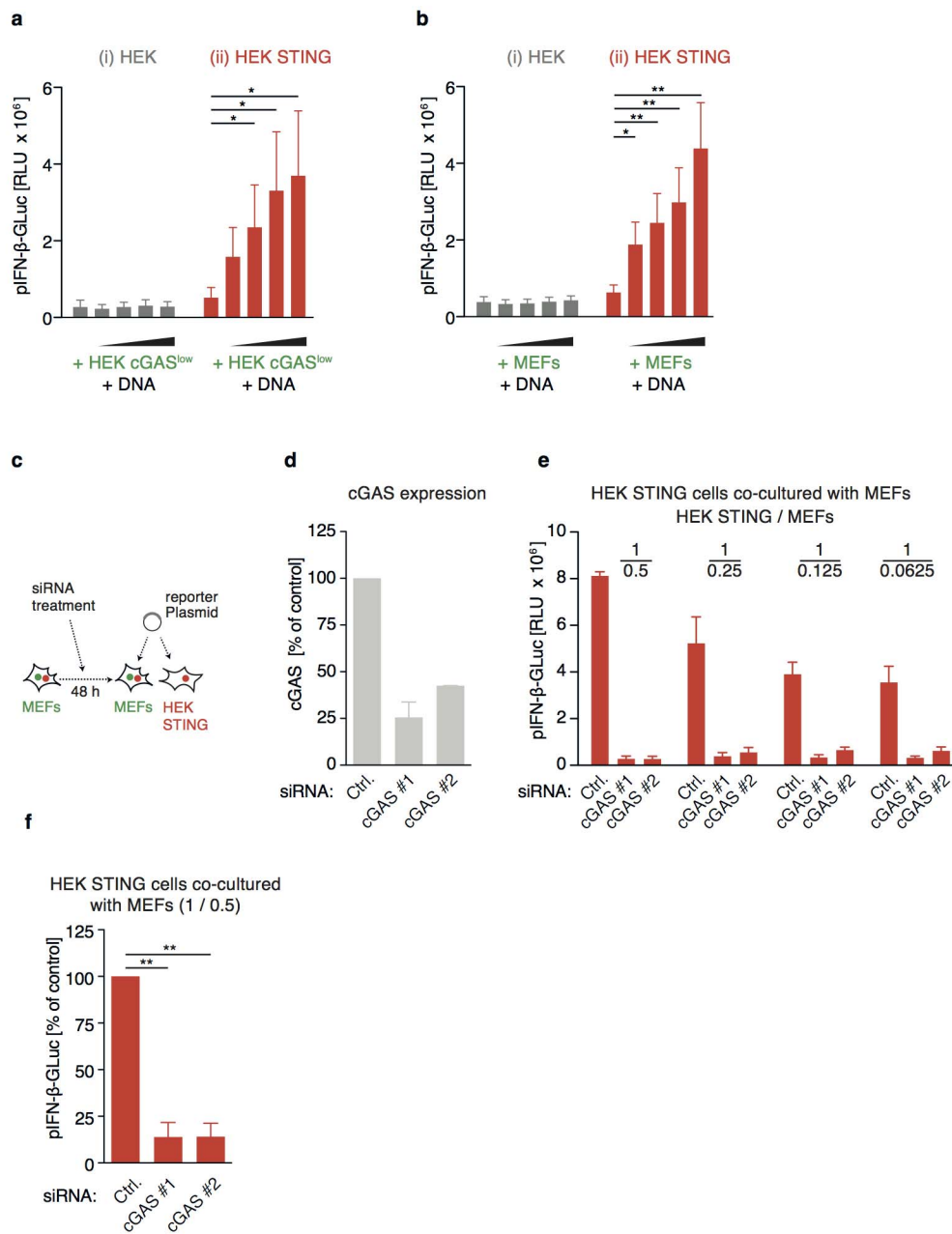
Extended Data Figure 2 | DNA-triggered cGAS activation induces IFN-β expression in adjacent cells via STING. a, **b**, HEK cells and HEK STING cells were co-cultured with increasing amounts of HEK cGAS^low cells (**a**) or primary MEFs (**b**) as depicted in Fig. 2c, d (ratio of HEK/HEK cGAS^low/MEFs was titrated ranging from 1:0.125 to 1:0.0156). Co-cultures were transfected with pIFN-β-GLuc and after 20 h transactivation of the reporter construct was assessed. Mean and s.e.m. of six experiments (**a**) or eight experiments (**b**) are depicted (*P < 0.05, **P < 0.01). **c**, Schematic view of the experimental set-up is shown: primary MEFs were silenced for cGAS expression using two independent siRNAs targeting cGAS or a control siRNA. Forty-eight hours later MEFs were co-cultured with HEK STING cells and then transfected with pIFN-β-GLuc and after an additional period of twenty hours transactivation of the reporter construct was assessed. **d**, cGAS expression in MEFs treated as in **c** was analysed by qPCR (data normalized to control siRNA condition). Mean values and s.e.m. of two independent experiments are depicted. **e**, Mean values and s.e.m. of duplicate measurements of one representative experiment, in which the ratio of HEK STING cells over MEFs was titrated ranging from 1:0.5 to 1:0.0625 is depicted. **f**, Mean values and s.e.m. (data normalized to control siRNA condition) of three independent experiments are depicted (HEK STING/MEF ratio is 1:0.5) (**P < 0.01).
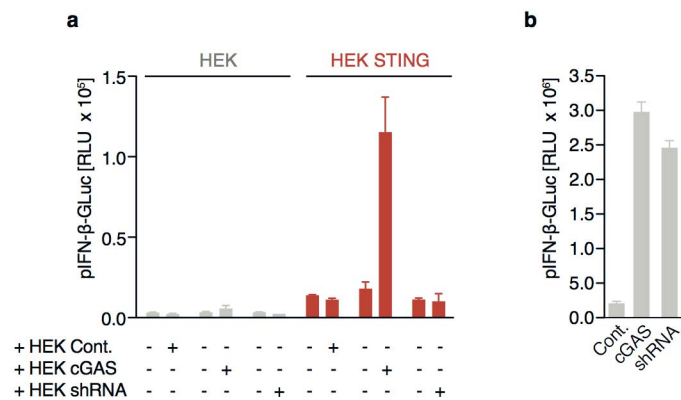
Extended Data Figure 3 | Overexpression of a RIG-I-stimulatory RNA molecule cannot confer activation of bystander cells. a, HEK cells were transfected with empty vector (Cont.), cGAS–GFP (cGAS) or a construct encoding a RIG-I-stimulatory shRNA molecule (shRNA). Twenty hours after transfection cells were collected, washed and added onto HEK cells or HEK STING cells expressing pIFN-β-GLuc. After 20 h of co-culture luciferase activity was measured. **b**, HEK STING cells were transfected as in **a** together with pIFN-β-GLuc and luciferase activity was measured 20 h after transfection. Mean and s.e.m. (biological duplicates) of one representative experiment out of two independent experiments are shown.
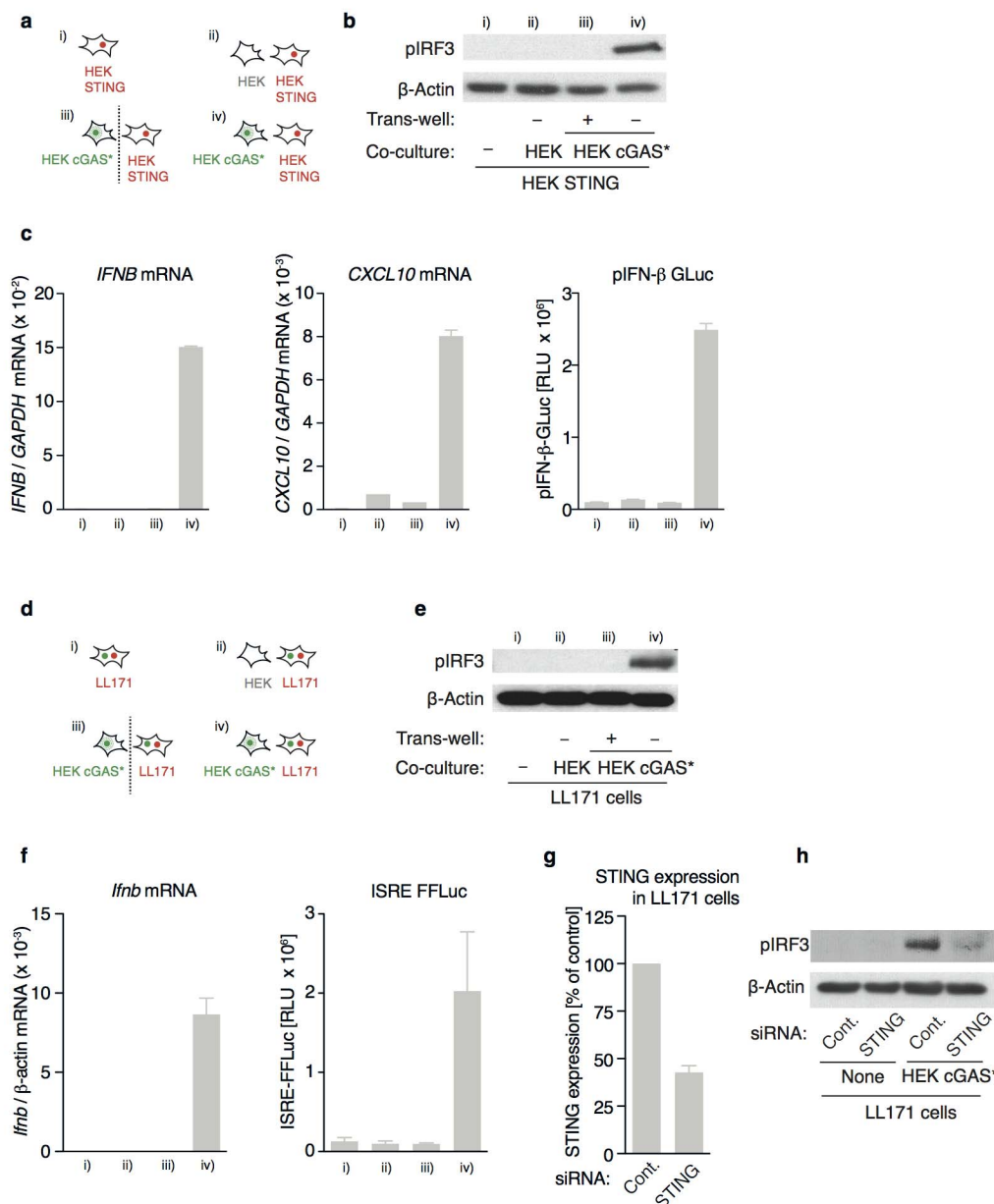
**Extended Data Figure 4 | cGAS-dependent bystander cell activation requires direct cell-to-cell contact. a, d,** Schematic view of the experimental set-up is depicted. HEK STING cells or LL171 cells were left untreated (i) or co-cultured with HEK cells (ii) or HEK cGAS* cells in the presence (iii) or absence (iv) of a trans-well system. **b, e,** After 4 h of co-culture, phosphorylation of IRF3 was determined in cellular lysates via immunoblotting. **c,** After 14 h, relative induction of *IFNB* and *CXCL10* in HEK STING cells was analysed via qPCR. In addition, HEK STING cells were transfected with pIFN-β-GLuc 20 h before donor cells were added. After 18 h luciferase activity in HEK STING cells was assessed. **f,** Relative induction of *Ifnb* in LL171 cells was determined via qPCR after 4 h. Furthermore, transactivation of an endogenous ISRE-reporter construct was assessed in LL171 cells after 14 h. **g,** LL171 cells were transfected with siRNA targeting STING or a control siRNA. Forty-eight hours after siRNA transfection relative expression of STING was determined by qPCR. Mean and s.e.m. of duplicate measurements of two independent experiments is shown. **h,** LL171 cells from **g** were co-cultured with HEK cGAS* cells and after 6 h phosphorylation of IRF3 was determined by immunoblotting. Mean and s.e.m. (biological duplicates) of one representative experiment out of two independent experiments are shown (**c, f**) or one representative experiment out of two independent experiments is shown (**b, e, h**).

a

LL171 cells

pIRF3

β-Actin

CBX: [μM]	0	0	100	150	200	0	100	150	200	0	100	150	200
	0		HEK				HEK cGAS*				CMA		

b

LL171 cells

ISRE-FFLuc [RLU x 10^5]

Control, HEK, HEK cGAS*, IFN-β

LL171 cells

ISRE-FFLuc [% of max]

LL171 cells + HEK cGAS*
LL171 cells + IFN-β

CBX: 0 ▬ 0 ▬
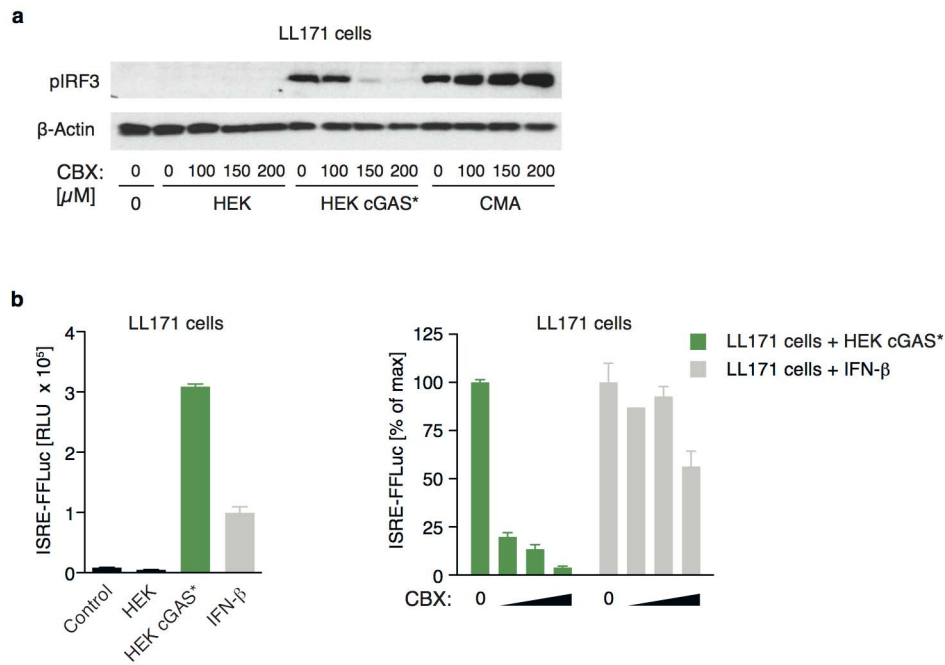
Extended Data Figure 5 | Carbenoxolone inhibits bystander effect in LL171 cells. a, b, LL171 cells were pre-treated with CBX (100 μM, 150 μM and 200 μM) 3 h before addition of HEK cells or HEK cGAS* cells. In addition, LL171 cells were stimulated with CMA (**a**) or recombinant IFN-α (250 U ml^{-1}). Phosphorylation of IRF3 (**a**) and luciferase activity of an endogenous ISRE-reporter construct (**b**) was determined in the cellular lysate 4 h and 14 h after stimulation, respectively. Mean and s.e.m. (biological duplicates) of one representative experiment out of two independent experiments are shown.
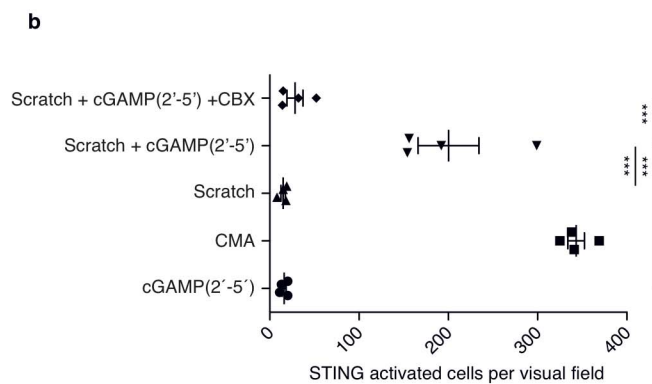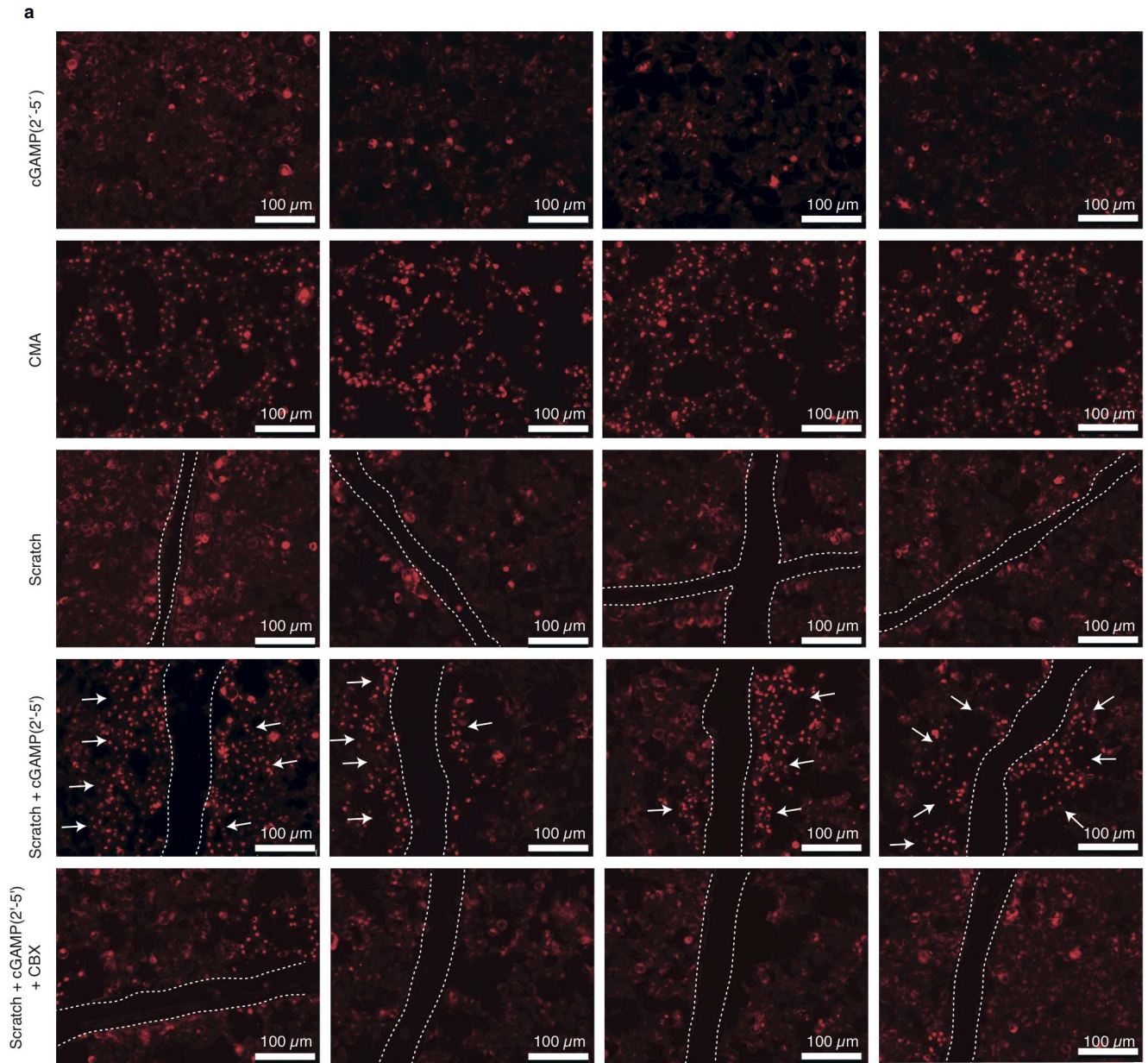
Extended Data Figure 6 | Scrape loading assays reveal a direct transfer of cGAMP(2′-5′) through gap junctions. a, b, HEK STING cells (STING in red) were either incubated with cGAMP(2′-5′), CMA or scratched in the presence of cGAMP(2′-5′). The latter condition was also performed in the presence of 150 μM CBX. STING activation was visualized 8 h later, whereas dashed lines follow the scratch margins and arrows highlight areas of STING complex assembly. Representative images of four independent experiments are shown (**a**) and STING-activated cells were quantified and depicted in a scatter plot (**b**). ***$P < 0.001$.

a

human GJA1 (Connexin 43) genomic locus

exon 2

UTR
CDS

```
GGTATTGAAGAGCATGGTAAGGTGAAAATGCGAGGGGGGTTGC          genomic sequence
          ATGGTAAGGTGAAAATGCGAGGG                    CRISPR target site
                               PAM
```

b

CX43/45^WT cell line 1

```
GGTATTGAAGAGCATGGTAAGGTGAAAATGCGAGGGGGGTTGC    allele 1 (wild type)
```

CX43/45^WT cell line 2

```
GGTATTGAAGAGCATGGTAAGGTGAAAATGCGAGGGGGGTTGC    allele 1 (wild type)
```

CX43/45^DKO cell line 1

```
GGTATTGAAGAGCATG...........................C    allele 1 (-26 bp)
GGTATTGAAGAGCATGGTAAGG..............GGGGTTGC    allele 2 (-13 bp)
```

CX43/45^DKO cell line 2

```
GGTATTGAAGAGCATGGTA............GAGGGGGTTGC    allele 1 (-12 bp)
complex                                        allele 2 (net +5 bp)
complex                                        allele 3 (net +2 bp)
```

c

human GJC1 (Connexin 45) genomic locus

exon 3

```
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT          genomic sequence
      CTTCCTGACTCGCCTGCTAGAGG                       CRISPR target site
                          PAM
```

d

CX43/45^WT cell line 1

```
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT    allele 1 (wild type)
                              A
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT    allele 2 (+1 bp)
```

CX43/45^WT cell line 2

```
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT    allele 1 (wild type)
CCATGAGTTGGAGCTTCCTGACTCGCCTG..AGGA.AGATT    allele 2 (-3 bp)
CCATGAGTTGGAGCTTCCTGACT........AGAGGAGATT    allele 3 (-8 bp)
```

CX43/45^DKO cell line 1

```
                              A
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT    allele 1 (+1 bp)
                              :::
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT    allele 2 (+417 bp)
```

CX43/45^DKO cell line 2

```
                                A
CCATGAGTTGGAGCTTCCTGACTCGCCTGAGTCAGGAGATT    allele 1 (+1 bp)
                              C
CCATGAGTTGGAGCTTCCTGACTCGCCTGCTAGAGGAGATT    allele 2 (+1 bp)
TCATGATTCGGAGCTTCCT.................GATT    allele 3 (-18 bp)
CCATGAGTTGGAG.......................ATT    allele 4 (-25 bp)
```

e

Extended Data Figure 7 | Deep sequencing results of CX43/CX45-targeted HEK STING cells generated by CRISPR/Cas9-mediated genome editing and western blot analysis of HEK STING CX43/45^DKO cells. a, c, For generating HEK STING CX43/45^DKO cells, a targeting strategy was devised based on hybrid gRNA sequences targeting Cas9 to the first coding exons of the respective genes. The open reading frame of CX43 (**a**) and CX45 (**c**) are delineated in red. PAM, protospacer adjacent motif. **b, d,** Deep-sequencing-based allele calls of targeted HEK STING cell lines as well as control cell lines are presented. Mutations are indicated in red letters, whereas the numbers in brackets indicate the net frame shifts. **e,** HEK STING CX43/45^WT cells and HEK STING CX43/45^DKO cells were analysed for CX43 and CX45 expression via immunoblotting. Of note, HEK STING CX43/45^DKO cell line 2 harbours an in-frame deletion for CX43 (−12 bp) and for CX45 (−18 bp), which probably accounts for the faint signal observed in the immunoblot (asterisk). Data are representative of three independent experiments.

a

b

HEK STING CX43/45^DKO
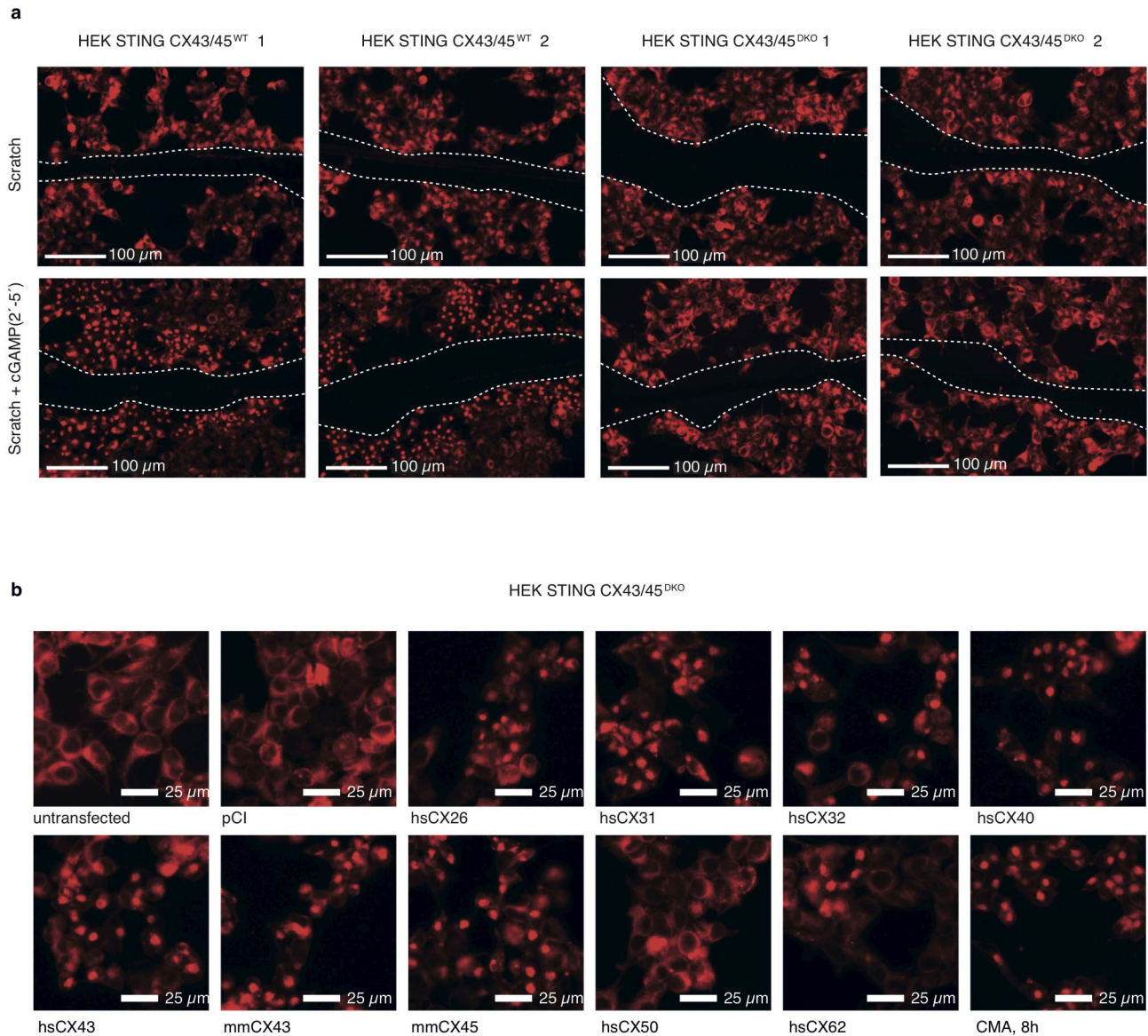
Extended Data Figure 8 | Scrape loading of cGAMP(2′-5′) into HEK STING CX43/45^WT and CX43/45^DKO cells and overexpression of distinct connexin members in HEK STING CX43/45^DKO cells. **a**, Fluorescence images of HEK STING CX43/45^WT and CX43/45^DKO cells (STING in red) wounded and overlaid with cGAMP(2′-5′). Wounded cells without addition of cGAMP(2′-5′) served as controls. Dashed line outlines the scratch margins. Representative images of $n = 2$ experiments are shown. **b**, Fluorescence images of HEK STING CX43/45^DKO co-cultured with HEK cGAS* and transfected with empty vector (pCI) and distinct members of human or murine connexins as indicated are depicted (pCI, CMA and mmCx45 as depicted in Fig. 4 are shown). Multimerization of STING was visualized 20 h after transfection. CMA stimulation for 8 h served as positive control. Representative images of $n = 2$ experiments are depicted.
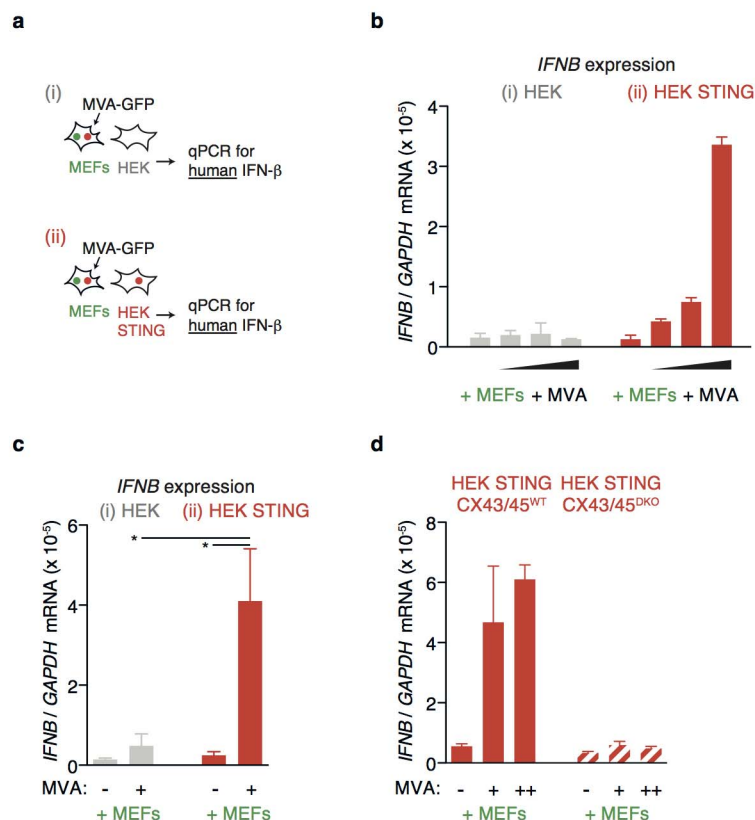
Extended Data Figure 9 | MVA-infected MEFs activate HEK STING cells *in trans* in a gap-junction dependent fashion. a, Schematic view of the experimental set-up for **b, c**: MEFs were infected with MVA–GFP for 3 h, washed three times and then loaded onto HEK cells or HEK STING cells that were then incubated overnight. Subsequently, human IFN-β expression was analysed by qPCR. **b, c,** A representative experiment with a titration of MVA–GFP (1.6×10^6, 0.8×10^6 and 0.16×10^6 virus particles per ml) is depicted

(**b**) and mean values and s.e.m. of three independent experiments at a concentration of 1.6×10^6 virus particles/ml are shown (**c**). **d,** Experiments were conducted as in **b**, now using HEK STING CX43/CX45^WT and HEK STING CX43/CX45^DKO cell lines as recipient cells. One representative experiment out of two independent experiments using 3.2×10^6 and 1.6×10^6 virus particles per ml is depicted in **d**.
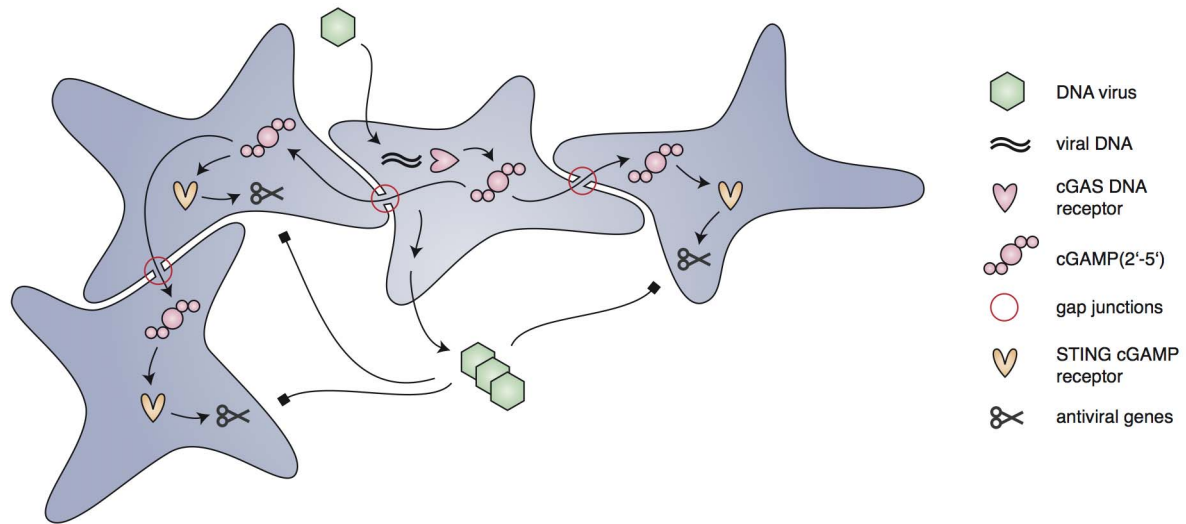
Extended Data Figure 10 | Schematic model of the mechanism of gap-junction-mediated local immune collaboration. On infection with a DNA virus, a cell senses the presence of cytosolic viral DNA through the receptor cGAS, thus activating the synthesis of the second messenger cGAMP(2′-5′). cGAMP(2′-5′) can pass through gap junctions into the cytosol of neighbouring cells, where it is detected by STING. The subsequent induction of an antiviral transcriptional program thus protects bystander cells from viral infection after the virus has successfully replicated in the cell initially infected.

LETTER

Isolation and characterization of a bat SARS–like coronavirus that uses the ACE2 receptor

Xing-Yi Ge[1]*, Jia-Lu Li[1]*, Xing-Lou Yang[1]*, Aleksei A. Chmura[2], Guangjian Zhu[2], Jonathan H. Epstein[2], Jonna K. Mazet[3], Ben Hu[1], Wei Zhang[1], Cheng Peng[1], Yu-Ji Zhang[1], Chu-Ming Luo[1], Bing Tan[1], Ning Wang[1], Yan Zhu[1], Gary Crameri[4], Shu-Yi Zhang[5], Lin-Fa Wang[4,6], Peter Daszak[2] & Zheng-Li Shi[1]

The 2002–3 pandemic caused by severe acute respiratory syndrome coronavirus (SARS-CoV) was one of the most significant public health events in recent history[1]. An ongoing outbreak of Middle East respiratory syndrome coronavirus[2] suggests that this group of viruses remains a key threat and that their distribution is wider than previously recognized. Although bats have been suggested to be the natural reservoirs of both viruses[3–5], attempts to isolate the progenitor virus of SARS-CoV from bats have been unsuccessful. Diverse SARS-like coronaviruses (SL-CoVs) have now been reported from bats in China, Europe and Africa[5–8], but none is considered a direct progenitor of SARS-CoV because of their phylogenetic disparity from this virus and the inability of their spike proteins to use the SARS-CoV cellular receptor molecule, the human angiotensin converting enzyme II (ACE2)[9,10]. Here we report whole-genome sequences of two novel bat coronaviruses from Chinese horseshoe bats (family: Rhinolophidae) in Yunnan, China: RsSHC014 and Rs3367. These viruses are far more closely related to SARS-CoV than any previously identified bat coronaviruses, particularly in the receptor binding domain of the spike protein. Most importantly, we report the first recorded isolation of a live SL-CoV (bat SL-CoV-WIV1) from bat faecal samples in Vero E6 cells, which has typical coronavirus morphology, 99.9% sequence identity to Rs3367 and uses ACE2 from humans, civets and Chinese horseshoe bats for cell entry. Preliminary _in vitro_ testing indicates that WIV1 also has a broad species tropism. Our results provide the strongest evidence to date that Chinese horseshoe bats are natural reservoirs of SARS-CoV, and that intermediate hosts may not be necessary for direct human infection by some bat SL-CoVs. They also highlight the importance of pathogen-discovery programs targeting high-risk wildlife groups in emerging disease hotspots as a strategy for pandemic preparedness.

The 2002–3 pandemic of SARS[1] and the ongoing emergence of the Middle East respiratory syndrome coronavirus (MERS-CoV)[2] demonstrate that CoVs are a significant public health threat. SARS-CoV was shown to use the human ACE2 molecule as its entry receptor, and this is considered a hallmark of its cross-species transmissibility[11]. The receptor binding domain (RBD) located in the amino-terminal region (amino acids 318–510) of the SARS-CoV spike (S) protein is directly involved in binding to ACE2 (ref. 12). However, despite phylogenetic evidence that SARS-CoV evolved from bat SL-CoVs, all previously identified SL-CoVs have major sequence differences from SARS-CoV in the RBD of their S proteins, including one or two deletions[6,9]. Replacing the RBD of one SL-CoV S protein with SARS-CoV S conferred the ability to use human ACE2 and replicate efficiently in mice[9,13]. However, to date, no SL-CoVs have been isolated from bats, and no wild-type SL-CoV of bat origin has been shown to use ACE2.

We conducted a 12-month longitudinal survey (April 2011–September 2012) of SL-CoVs in a colony of _Rhinolophus sinicus_ at a single location in Kunming, Yunnan Province, China (Extended Data Table 1). A total of 117 anal swabs or faecal samples were collected from individual bats using a previously published method[5,14]. A one-step reverse transcription (RT)-nested PCR was conducted to amplify the RNA-dependent RNA polymerase (RdRP) motifs A and C, which are conserved among alphacoronaviruses and betacoronaviruses[15].

Twenty-seven of the 117 samples (23%) were classed as positive by PCR and subsequently confirmed by sequencing. The species origin of all positive samples was confirmed to be _R. sinicus_ by cytochrome b sequence analysis, as described previously[16]. A higher prevalence was observed in samples collected in October (30% in 2011 and 48.7% in 2012) than those in April (7.1% in 2011) or May (7.4% in 2012) (Extended Data Table 1). Analysis of the S protein RBD sequences indicated the presence of seven different strains of SL-CoVs (Fig. 1a and Extended Data Figs 1 and 2). In addition to RBD sequences, which closely matched previously described SL-CoVs (Rs672, Rf1 and HKU3)[5,8,17,18], two novel strains (designated SL-CoV RsSHC014 and Rs3367) were discovered. Their full-length genome sequences were determined, and both were found to be 29,787 base pairs in size (excluding the poly(A) tail). The overall nucleotide sequence identity of these two genomes with human SARS-CoV (Tor2 strain) is 95%, higher than that observed previously for bat SL-CoVs in China (88–92%)[5,8,17,18] or Europe (76%)[6] (Extended Data Table 2 and Extended Data Figs 3 and 4). Higher sequence identities were observed at the protein level between these new SL-CoVs and SARS-CoVs (Extended Data Tables 3 and 4). To understand the evolutionary origin of these two novel SL-CoV strains, we conducted recombination analysis with the Recombination Detection Program 4.0 package[19] using available genome sequences of bat SL-CoV strains (Rf1, Rp3, Rs672, Rm1, HKU3 and BM48-31) and human and civet representative SARS-CoV strains (BJ01, SZ3, Tor2 and GZ02). Three breakpoints were detected with strong _P_ values ($<10^{-20}$) and supported by similarity plot and bootscan analysis (Extended Data Fig. 5a, b). Breakpoints were located at nucleotides 20,827, 26,553 and 28,685 in the Rs3367 (and RsSHC014) genome, and generated recombination fragments covering nucleotides 20,827–26,533 (5,727 nucleotides) (including partial open reading frame (ORF) 1b, full-length S, ORF3, E and partial M gene) and nucleotides 26,534–28,685 (2,133 nucleotides) (including partial ORF M, full-length ORF6, ORF7, ORF8 and partial N gene). Phylogenetic analysis using the major and minor parental regions suggested that Rs3367, or RsSHC014, is the descendent of a recombination of lineages that ultimately lead to SARS-CoV and SL-CoV Rs672 (Fig. 1b).

The most notable sequence differences between these two new SL-CoVs and previously identified SL-CoVs is in the RBD regions of their S proteins. First, they have higher amino acid sequence identity to SARS-CoV (85% and 96% for RsSHC014 and Rs3367, respectively). Second, there are no deletions and they have perfect sequence alignment with the SARS-CoV RBD region (Extended Data Figs 1 and 2). Structural
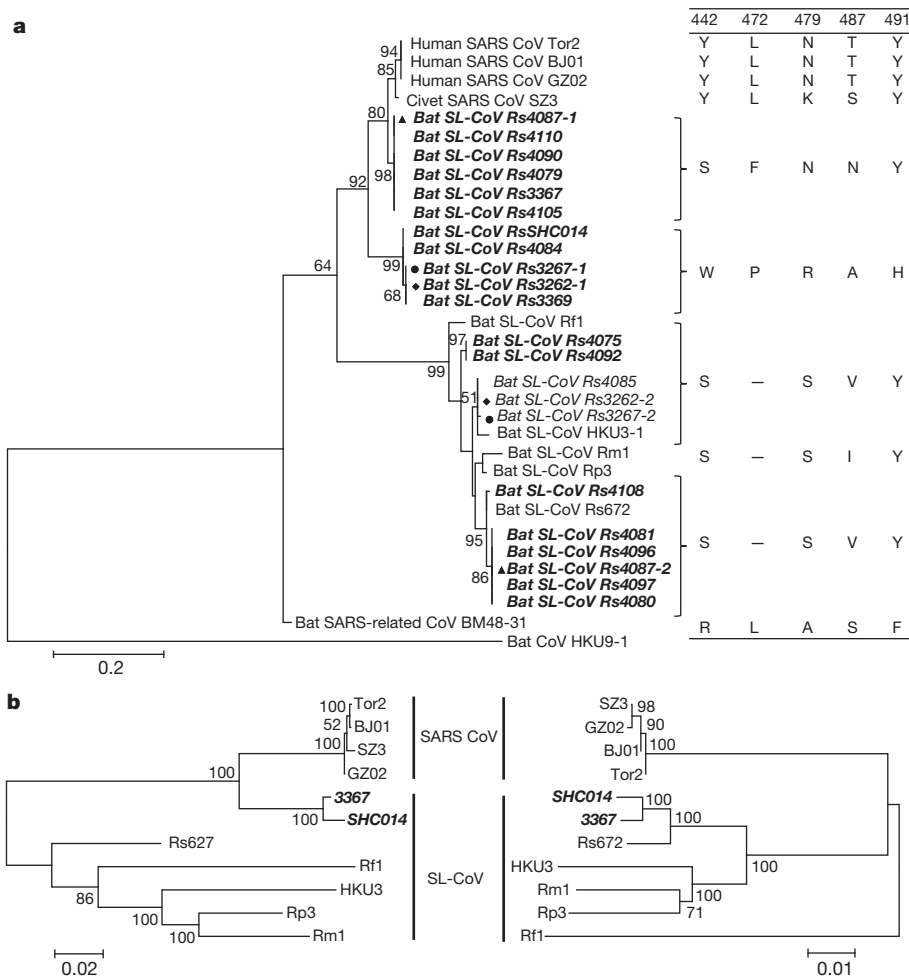

Figure 1 | Phylogenetic tree based on amino acid sequences of the S RBD region and the two parental regions of bat SL-CoV Rs3367 or RsSHC014. **a**, SARS-CoV S protein amino acid residues 310–520 were aligned with homologous regions of bat SL-CoVs using the ClustalW software. A maximum-likelihood phylogenetic tree was constructed using a Poisson model with bootstrap values determined by 1,000 replicates in the MEGA5 software package. The RBD sequences identified in this study are in bold and named by the sample numbers. The key amino acid residues involved in interacting with the human ACE2 molecule are indicated on the right of the tree. SARS-CoV GZ02, BJ01 and Tor2 were isolated from patients in the early, middle and late phase, respectively, of the SARS outbreak in 2003. SARS-CoV SZ3 was identified from *Paguma larvata* in 2003 collected in Guangdong, China. SL-CoV Rp3, Rs672 and HKU3-1 were identified from *R. sinicus* collected in China (respectively: Guangxi, 2004; Guizhou, 2006; Hong Kong, 2005). Rf1 and Rm1 were identified from

R. ferrumequinum and *R. macrotis*, respectively, collected in Hubei, China, in 2004. Bat SARS-related CoV BM48-31 was identified from *R. blasii* collected in Bulgaria in 2008. Bat CoV HKU9-1 was identified from *Rousettus leschenaultii* collected in Guangdong, China in 2005/2006 and used as an outgroup. All sequences in bold and italics were identified in the current study. Filled triangles, circles and diamonds indicate samples with co-infection by two different SL-CoVs. '–' indicates the amino acid deletion. **b**, Phylogenetic origins of the two parental regions of Rs3367 or RsSHC014. Maximum likelihood phylogenetic trees were constructed from alignments of two fragments covering nucleotides 20,827–26,533 (5,727 nucleotides) and 26,534 –28,685 (2,133 nucleotides) of the Rs3367 genome, respectively. For display purposes, the trees were midpoint rooted. The taxa were annotated according to strain names: SARS-CoV, SARS coronavirus; SARS-like CoV, bat SARS-like coronavirus. The two novel SL-CoVs, Rs3367 and RsSHC014, are in bold and italics.

and mutagenesis studies have previously identified five key residues (amino acids 442, 472, 479, 487 and 491) in the RBD of the SARS-CoV S protein that have a pivotal role in receptor binding[20,21]. Although all five residues in the RsSHC014 S protein were found to be different from those of SARS-CoV, two of the five residues in the Rs3367 RBD were conserved (Fig. 1 and Extended Data Fig. 1).

Despite the rapid accumulation of bat CoV sequences in the last decade, there has been no report of successful virus isolation[6,22,23]. We attempted isolation from SL-CoV PCR-positive samples. Using an optimized protocol and Vero E6 cells, we obtained one isolate which caused cytopathic effect during the second blind passage. Purified virions displayed typical coronavirus morphology under electron microscopy (Fig. 2). Sequence analysis using a sequence-independent amplification method[14] to avoid PCR-introduced contamination indicated that the isolate was almost identical to Rs3367, with 99.9% nucleotide genome sequence identity and 100% amino acid sequence identity for the S1 region. The new isolate was named SL-CoV-WIV1.

To determine whether WIV1 can use ACE2 as a cellular entry receptor, we conducted virus infectivity studies using HeLa cells expressing or not expressing ACE2 from humans, civets or Chinese horseshoe bats. We found that WIV1 is able to use ACE2 of different origins as an entry receptor and replicated efficiently in the ACE2-expressing cells (Fig. 3). This is, to our knowledge, the first identification of a wild-type bat SL-CoV capable of using ACE2 as an entry receptor.

To assess its cross-species transmission potential, we conducted infectivity assays in cell lines from a range of species. Our results (Fig. 4 and Extended Data Table 5) indicate that bat SL-CoV-WIV1 can grow in human alveolar basal epithelial (A549), pig kidney 15 (PK-15) and *Rhinolophus sinicus* kidney (RSKT) cell lines, but not in human cervix (HeLa), Syrian golden hamster kidney (BHK21), *Myotis davidii* kidney (BK), *Myotis chinensis* kidney (MCKT), *Rousettus leschenaulti* kidney (RLK) or *Pteropus alecto* kidney (PaKi) cell lines. Real-time RT–PCR indicated that WIV1 replicated much less efficiently in A549, PK-15 and RSKT cells than in Vero E6 cells (Fig. 4).
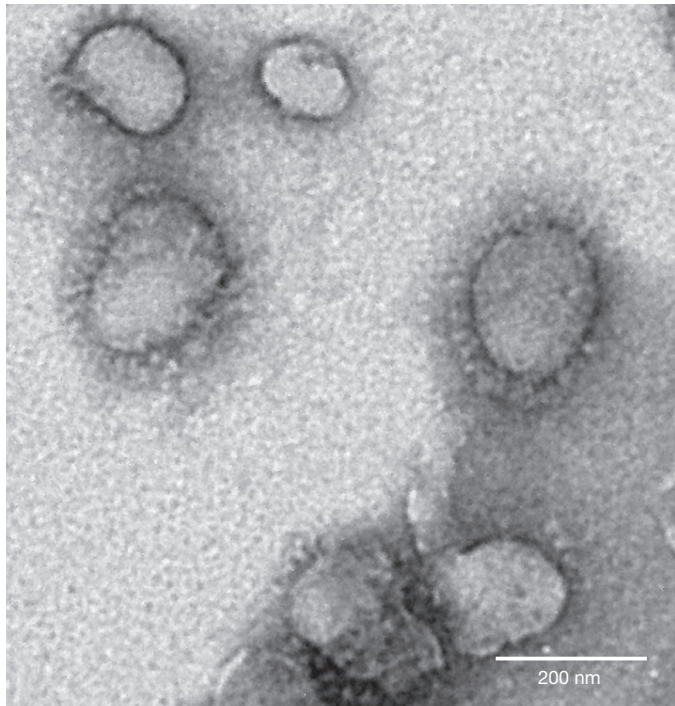
Figure 2 | Electron micrograph of purified virions. Virions from a 10-ml culture were collected, fixed and concentrated/purified by sucrose gradient centrifugation. The pelleted viral particles were suspended in 100 μl PBS, stained with 2% phosphotungstic acid (pH 7.0) and examined directly using a Tecnai transmission electron microscope (FEI) at 200 kV.

To assess the cross-neutralization activity of human SARS-CoV sera against WIV1, we conducted serum-neutralization assays using nine convalescent sera from SARS patients collected in 2003. The results showed that seven of these were able to completely neutralize 100 tissue culture infectious dose 50 ($TCID_{50}$) WIV1 at dilutions of 1:10 to 1:40, further confirming the close relationship between WIV1 and SARS-CoV.

Our findings have important implications for public health. First, they provide the clearest evidence yet that SARS-CoV originated in bats. Our previous work provided phylogenetic evidence of this[5], but the lack of an isolate or evidence that bat SL-CoVs can naturally infect human cells, until now, had cast doubt on this hypothesis. Second, the lack of capacity of SL-CoVs to use of ACE2 receptors has previously been considered as the key barrier for their direct spillover into humans, supporting the suggestion that civets were intermediate hosts for SARS-CoV adaptation to human transmission during the SARS outbreak[24]. However, the ability of SL-CoV-WIV1 to use human ACE2 argues against the necessity of this step for SL-CoV-WIV1 and suggests that direct bat-to-human infection is a plausible scenario for some bat SL-CoVs. This has implications for public health control measures in the face of potential spillover of a diverse and growing pool of recently discovered SARS-like CoVs with a wide geographic distribution.

Our findings suggest that the diversity of bat CoVs is substantially higher than that previously reported. In this study we were able to demonstrate the circulation of at least seven different strains of SL-CoVs within a single colony of *R. sinicus* during a 12-month period. The high genetic diversity of SL-CoVs within this colony was mirrored by high phenotypic diversity in the differential use of ACE2 by different strains. It would therefore not be surprising if further surveillance reveals a broad diversity of bat SL-CoVs that are able to use ACE2, some of which may have even closer homology to SARS-CoV than SL-CoV-WIV1. Our results—in addition to the recent demonstration of MERS-CoV in a Saudi Arabian bat[25], and of bat CoVs closely related to MERS-CoV in China, Africa, Europe and North America[3,26,27]—suggest that bat coronaviruses remain a substantial global threat to public health.

Finally, this study demonstrates the public health importance of pathogen discovery programs targeting wildlife that aim to identify the 'known unknowns'—previously unknown viral strains closely related to known pathogens. These programs, focused on specific high-risk wildlife groups and hotspots of disease emergence, may be a critical part of future global strategies to predict, prepare for, and prevent pandemic emergence[28].
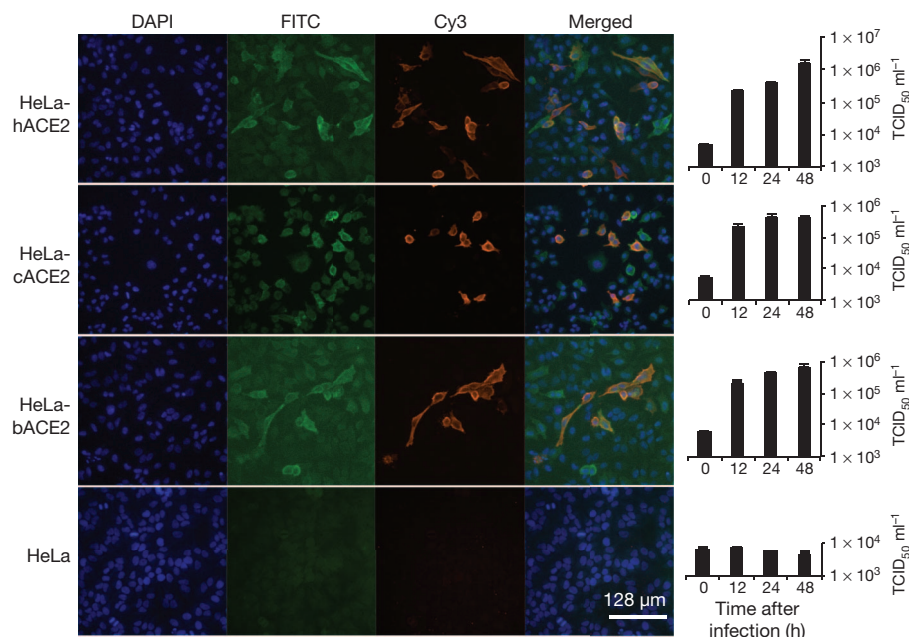
Figure 3 | Analysis of receptor usage of SL-CoV-WIV1 determined by immunofluorescence assay and real-time PCR. Determination of virus infectivity in HeLa cells with and without the expression of ACE2. b, bat; c, civet; h, human. ACE2 expression was detected with goat anti-humanACE2 antibody followed by fluorescein isothiocyanate (FITC)-conjugated donkey anti-goat IgG. Virus replication was detected with rabbit antibody against the SL-CoV Rp3 nucleocapsid protein followed by cyanine 3 (Cy3)-conjugated mouse anti-rabbit IgG. Nuclei were stained with DAPI (4′,6-diamidino-2-phenylindole). The columns (from left to right) show staining of nuclei (blue), ACE2 expression (green), virus replication (red), merged triple-stained images and real-time PCR results, respectively. ($n = 3$); error bars represent standard deviation.
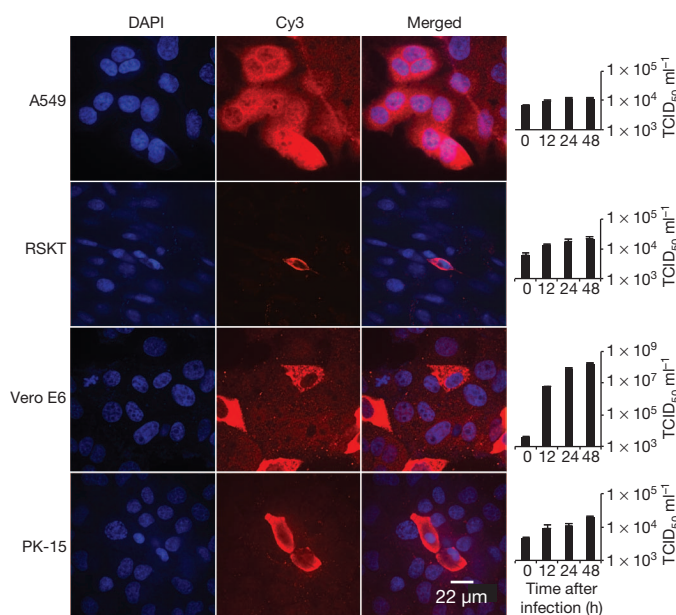
Figure 4 | Analysis of host range of SL-CoV-WIV1 determined by immunofluorescence assay and real-time PCR. Virus infection in A549, RSKT, Vero E6 and PK-15 cells. Virus replication was detected as described for Fig. 3. The columns (from left to right) show staining of nuclei (blue), virus replication (red), merged double-stained images and real-time PCR results, respectively. $n = 3$; error bars represent s.d.

METHODS SUMMARY

Throat and faecal swabs or fresh faecal samples were collected in viral transport medium as described previously[14]. All PCR was conducted with the One-Step RT–PCR kit (Invitrogen). Primers targeting the highly conserved regions of the RdRP gene were used for detection of all alphacoronaviruses and betacoronaviruses as described previously[15]. Degenerate primers were designed on the basis of all available genomic sequences of SARS-CoVs and SL-CoVs and used for amplification of the RBD sequences of S genes or full-length genomic sequences. Degenerate primers were used for amplification of the bat *ACE2* gene as described previously[29]. PCR products were gel purified and cloned into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence. PCR-positive faecal samples (in 200 µl buffer) were gradient centrifuged at 3,000–12,000g and supernatant diluted at 1:10 in DMEM before being added to Vero E6 cells. After incubation at 37 °C for 1 h, inocula were removed and replaced with fresh DMEM with 2% FCS. Cells were incubated at 37 °C and checked daily for cytopathic effect. Cell lines from different origins were grown on coverslips in 24-well plates and inoculated with the novel SL-CoV at a multiplicity of infection of 10. Virus replication was detected at 24 h after infection using rabbit antibodies against the SL-CoV Rp3 nucleocapsid protein followed by Cy3-conjugated goat anti-rabbit IgG.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 May; accepted 18 September 2013.
Published online 30 October 2013.

1. Ksiazek, T. G. *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348,** 1953–1966 (2003).
2. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* **367,** 1814–1820 (2012).
3. Anthony, S. J. *et al.* Coronaviruses in bats from Mexico. *J. Gen. Virol.* **94,** 1028–1038 (2013).
4. Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* **495,** 251–254 (2013).
5. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310,** 676–679 (2005).
6. Drexler, J. F. *et al.* Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J. Virol.* **84,** 11336–11349 (2010).
7. Tong, S. *et al.* Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg. Infect. Dis.* **15,** 482–485 (2009).
8. Lau, S. K. P. *et al.* Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl Acad. Sci. USA* **102,** 14040–14045 (2005).
9. Ren, W. *et al.* Difference in receptor usage between severe acute respiratory syndrome (SARS) coronavirus and SARS-like coronavirus of bat origin. *J. Virol.* **82,** 1899–1907 (2008).
10. Hon, C. C. *et al.* Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J. Virol.* **82,** 1819–1826 (2008).
11. Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426,** 450–454 (2003).
12. Wong, S. K., Li, W., Moore, M. J., Choe, H. & Farzan, M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J. Biol. Chem.* **279,** 3197–3201 (2004).
13. Becker, M. M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl Acad. Sci. USA* **105,** 19944–19949 (2008).
14. Li, Y. *et al.* Host range, prevalence, and genetic diversity of adenoviruses in bats. *J. Virol.* **84,** 3889–3897 (2010).
15. De Souza Luna, L. K. *et al.* Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *J. Clin. Microbiol.* **45,** 1049–1052 (2007).
16. Cui, J. *et al.* Evolutionary relationships between bat coronaviruses and their hosts. *Emerg. Infect. Dis.* **13,** 1526–1532 (2007).
17. Yuan, J. *et al.* Intraspecies diversity of SARS-like coronaviruses in *Rhinolophus sinicus* and its implications for the origin of SARS coronaviruses in humans. *J. Gen. Virol.* **91,** 1058–1062 (2010).
18. Ren, W. *et al.* Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J. Gen. Virol.* **87,** 3355–3359 (2006).
19. Martin, D. P. *et al.* RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26,** 2462–2463 (2010).
20. Wu, K., Peng, G., Wilken, M., Geraghty, R. J. & Li, F. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *J. Biol. Chem.* **287,** 8904–8911 (2012).
21. Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* **24,** 1634–1643 (2005).
22. Lau, S. K. *et al.* Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related *Rhinolophus* bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* **84,** 2808–2819 (2010).
23. Lau, S. K. *et al.* Coexistence of different genotypes in the same bat and serological characterization of *Rousettus* bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J. Virol.* **84,** 11385–11394 (2010).
24. Song, H. D. *et al.* Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl Acad. Sci. USA* **102,** 2430–2435 (2005).
25. Memish, Z. A. *et al.* Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* **19,** 11 (2013).
26. Chan, J. F. *et al.* Is the discovery of the novel human betacoronavirus 2c EMC/2012 (HCoV-EMC) the beginning of another SARS-like pandemic? *J. Infect.* **65,** 477–489 (2012).
27. Ithete, N. L. *et al.* Close relative of human Middle East respiratory syndrome coronavirus in bat, South Africa. *Emerg. Infect. Dis.* **19,** 1697–1699 (2013).
28. Morse, S. S. *et al.* Prediction and prevention of the next pandemic zoonosis. *Lancet* **380,** 1956–1965 (2012).
29. Hou, Y. *et al.* Angiotensin-converting enzyme 2 (ACE2) proteins of different bat species confer variable susceptibility to SARS-CoV entry. *Arch. Virol.* **155,** 1563–1569 (2010).

Author Contributions Z.-L.S. and P.D. designed and coordinated the study. X.-Y.G., J.-L.L. and X.-L.Y. conducted majority of experiments and contributed equally to the study. A.A.C., B.H., W.Z., C.P., Y.-J.Z., C.-M.L., B.T., N.W. and Y.Z. conducted parts of the experiments and analyses. J.H.E., J.K.M. and S.-Y.Z. coordinated the field study. X.-Y.G., J.-L.L., X.-L.Y., B.T. and G.-J.Z. collected the samples. G.C. and L.-F.W. designed and supervised part of the experiments. All authors contributed to the interpretations and conclusions presented. Z.-L.S. and X-Y.G. wrote the manuscript with significant contributions from P.D. and L.-F.W. and input from all authors.

METHODS

Sampling. Bats were trapped in their natural habitat as described previously[5]. Throat and faecal swab samples were collected in viral transport medium (VTM) composed of Hank's balanced salt solution, pH 7.4, containing BSA (1%), amphotericin (15 µg ml^{-1}), penicillin G (100 U ml^{-1}) and streptomycin (50 µg ml^{-1}). To collect fresh faecal samples, clean plastic sheets measuring 2.0 by 2.0 m were placed under known bat roosting sites at about 18:00 h each evening. Relatively fresh faecal samples were collected from sheets at approximately 05:30–06:00 the next morning and placed in VTM. Samples were transported to the laboratory and stored at −80 °C until use. All animals trapped for this study were released back to their habitat after sample collection. All sampling processes were performed by veterinarians with approval from Animal Ethics Committee of the Wuhan Institute of Virology (WIVH05210201) and EcoHealth Alliance under an inter-institutional agreement with University of California, Davis (UC Davis protocol no. 16048).

RNA extraction, PCR and sequencing. RNA was extracted from 140 µl of swab or faecal samples with a Viral RNA Mini Kit (Qiagen) following the manufacturer's instructions. RNA was eluted in 60 µl RNAse-free buffer (buffer AVE, Qiagen), then aliquoted and stored at −80 °C. One-step RT–PCR (Invitrogen) was used to detect coronavirus sequences as described previously[15]. First round PCR was conducted in a 25-µl reaction mix containing 12.5 µl PCR 2× reaction mix buffer, 10 pmol of each primer, 2.5 mM MgSO4, 20 U RNase inhibitor, 1 µl SuperScript III/ Platinum Taq Enzyme Mix and 5 µl RNA. Amplification of the RdRP-gene fragment was performed as follows: 50 °C for 30 min, 94 °C for 2 min, followed by 40 cycles consisting of 94 °C for 15 s, 62 °C for 15 s, 68 °C for 40 s, and a final extension of 68 °C for 5 min. Second round PCR was conducted in a 25-µl reaction mix containing 2.5 µl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl2, 0.5 mM dNTP, 0.1 µl Platinum Taq Enzyme (Invitrogen) and 1 µl first round PCR product. The amplification of RdRP-gene fragment was performed as follows: 94 °C for 5 min followed by 35 cycles consisting of 94 °C for 30 s, 52 °C for 30 s, 72 °C for 40 s, and a final extension of 72 °C for 5 min.

To amplify the RBD region, one-step RT–PCR was performed with primers designed based on available SARS-CoV or bat SL-CoVs (first round PCR primers; F, forward; R, reverse: CoVS931F-5′-VWGADGTTGTKAGRTTYCCT-3′ and CoVS1909R-5′-TAARACAVCCWGGCYTGWGT-3′; second PCR primers: CoVS 951F-5′-TGTKAGRTTYCCTAAYATTAC-3′ and CoVS1805R-5′-ACATCYTG ATANARAACAGC-3′). First-round PCR was conducted in a 25-µl reaction mix as described above except primers specific for the S gene were used. The amplification of the RBD region of the S gene was performed as follows: 50 °C for 30 min, 94 °C for 2 min, followed by 35 cycles consisting of 94 °C for 15 s, 43 °C for 15 s, 68 °C for 90 s, and a final extension of 68 °C for 5 min. Second-round PCR was conducted in a 25-µl reaction mix containing 2.5 µl PCR reaction buffer, 5 pmol of each primer, 50 mM MgCl2, 0.5 mM dNTP, 0.1 µl Platinum Taq Enzyme (Invitrogen) and 1 µl first round PCR product. Amplification was performed as follows: 94 °C for 5 min followed by 40 cycles consisting of 94 °C for 30 s, 41 °C for 30 s, 72 °C for 60 s, and a final extension of 72 °C for 5 min.

PCR products were gel purified and cloned into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence for each of the amplified regions.

Sequencing full-length genomes. Degenerate coronavirus primers were designed based on all available SARS-CoV and bat SL-CoV sequences in GenBank and specific primers were designed from genome sequences generated from previous rounds of sequencing in this study (primer sequences will be provided upon request). All PCRs were conducted using the One-Step RT–PCR kit (Invitrogen). The 5′ and 3′ genomic ends were determined using the 5′ or 3′ RACE kit (Roche), respectively. PCR products were gel purified and sequenced directly or following cloning into pGEM-T Easy Vector (Promega). At least four independent clones were sequenced to obtain a consensus sequence for each of the amplified regions and each region was sequenced at least twice.

Sequence analysis and databank accession numbers. Routine sequence management and analysis was carried out using DNAStar or Geneious. Sequence alignment and editing was conducted using ClustalW, BioEdit or GeneDoc. Maximum Likelihood phylogenetic trees based on the protein sequences were constructed using a Poisson model with bootstrap values determined by 1,000 replicates in the MEGA5 software package.

Sequences obtained in this study have been deposited in GenBank as follows (accession numbers given in parenthesis): full-length genome sequence of SL-CoV RsSHC014 and Rs3367 (KC881005, KC881006); full-length sequence of WIV1 S (KC881007); RBD (KC880984-KC881003); ACE2 (KC8810040). SARS-CoV sequences used in this study: human SARS-CoV strains Tor2 (AY274119), BJ01 (AY278488), GZ02 (AY390556) and civet SARS-CoV strain SZ3 (AY304486). Bat coronavirus sequences used in this study: Rs672 (FJ588686), Rp3 (DQ071615), Rf1 (DQ412042), Rm1 (DQ412043), HKU3-1 (DQ022305), BM48-31 (NC_014470), HKU9-1 (NC_009021), HKU4 (NC_009019), HKU5 (NC_009020), HKU8 (DQ249228),

HKU2 (EF203067), BtCoV512 (NC_009657), 1A (NC_010437). Other coronavirus sequences used in this study: HCoV-229E (AF304460), HCoV-OC43 (AY391777), HCoV-NL63 (AY567487), HKU1 (NC_006577), EMC (JX869059), FIPV (NC_002306), PRCV (DQ811787), BWCoV (NC_010646), MHV (AY700211), IBV (AY851295).

Amplification, cloning and expression of the bat ACE2 gene. Construction of expression clones for human and civet ACE2 in pcDNA3.1 has been described previously[29]. Bat ACE2 was amplified from a *R. sinicus* (sample no. 3357). In brief, total RNA was extracted from bat rectal tissue using the RNeasy Mini Kit (Qiagen). First-strand complementary DNA was synthesized from total RNA by reverse transcription with random hexamers. Full-length bat *ACE2* fragments were amplified using forward primer bAF2 and reverse primer bAR2 (ref. 29). The *ACE2* gene was cloned into pCDNA3.1 with KpnI and XhoI, and verified by sequencing. Purified *ACE2* plasmids were transfected to HeLa cells. After 24 h, lysates of HeLa cells expressing human, civet, or bat ACE2 were confirmed by western blot or immunofluorescence assay.

Western blot analysis. Lysates of cells or filtered supernatants containing pseudoviruses were separated by SDS–PAGE, followed by transfer to a nitrocellulose membrane (Millipore). For detection of S protein, the membrane was incubated with rabbit anti-Rp3 S fragment (amino acids 561–666) polyantibodies (1:200), and the bound antibodies were detected by alkaline phosphatase (AP)-conjugated goat anti-rabbit IgG (1:1,000). For detection of HIV-1 p24 in supernatants, monoclonal antibody against HIV p24 (p24 MAb) was used as the primary antibody at a dilution of 1:1,000, followed by incubation with AP-conjugated goat anti-mouse IgG at the same dilution. To detect the expression of ACE2 in HeLa cells, goat antibody against the human ACE2 ectodomain (1:500) was used as the first antibody, followed by incubation with horseradish peroxidase-conjugated donkey anti-goat IgG (1:1,000).

Virus isolation. Vero E6 cell monolayers were maintained in DMEM supplemented with 10% FCS. PCR-positive samples (in 200 µl buffer) were gradient centrifuged at 3,000–12,000g, and supernatant was diluted 1:10 in DMEM before being added to Vero E6 cells. After incubation at 37 °C for 1 h, inocula were removed and replaced with fresh DMEM with 2% FCS. Cells were incubated at 37 °C for 3 days and checked daily for cytopathic effect. Double-dose triple antibiotics penicillin/ streptomycin/amphotericin (Gibco) were included in all tissue culture media (penicillin 200 IU ml^{-1}, streptomycin 0.2 mg ml^{-1}, amphotericin 0.5 µg ml^{-1}). Three blind passages were carried out for each sample. After each passage, both the culture supernatant and cell pellet were examined for presence of virus by RT–PCR using primers targeting the RdRP or S gene. Virions in supernatant (10 ml) were collected and fixed using 0.1% formaldehyde for 4 h, then concentrated by ultracentrifugation through a 20% sucrose cushion (5 ml) at 80,000g for 90 min using a Ty90 rotor (Beckman). The pelleted viral particles were suspended in 100 µl PBS, stained with 2% phosphotungstic acid (pH 7.0) and examined using a Tecnai transmission electron microscope (FEI) at 200 kV.

Virus infectivity detected by immunofluorescence assay. Cell lines used for this study and their culture conditions are summarized in Extended Data Table 5. Virus titre was determined in Vero E6 cells by cytopathic effect (CPE) counts. Cell lines from different origins and HeLa cells expressing ACE2 from human, civet or Chinese horseshoe bat were grown on coverslips in 24-well plates (Corning) incubated with bat SL-CoV-WIV1 at a multiplicity of infection = 10 for 1 h. The inoculum was removed and washed twice with PBS and supplemented with medium. HeLa cells without ACE2 expression and Vero E6 cells were used as negative and positive controls, respectively. At 24 h after infection, cells were washed with PBS and fixed with 4% formaldehyde in PBS (pH 7.4) for 20 min at 4 °C. ACE2 expression was detected using goat anti-human ACE2 immunoglobulin (R&D Systems) followed by FITC-labelled donkey anti-goat immunoglobulin (PTGLab). Virus replication was detected using rabbit antibody against the SL-CoV Rp3 nucleocapsid protein followed by Cy3-conjugated mouse anti-rabbit IgG. Nuclei were stained with DAPI. Staining patterns were examined using a FV1200 confocal microscope (Olympus).
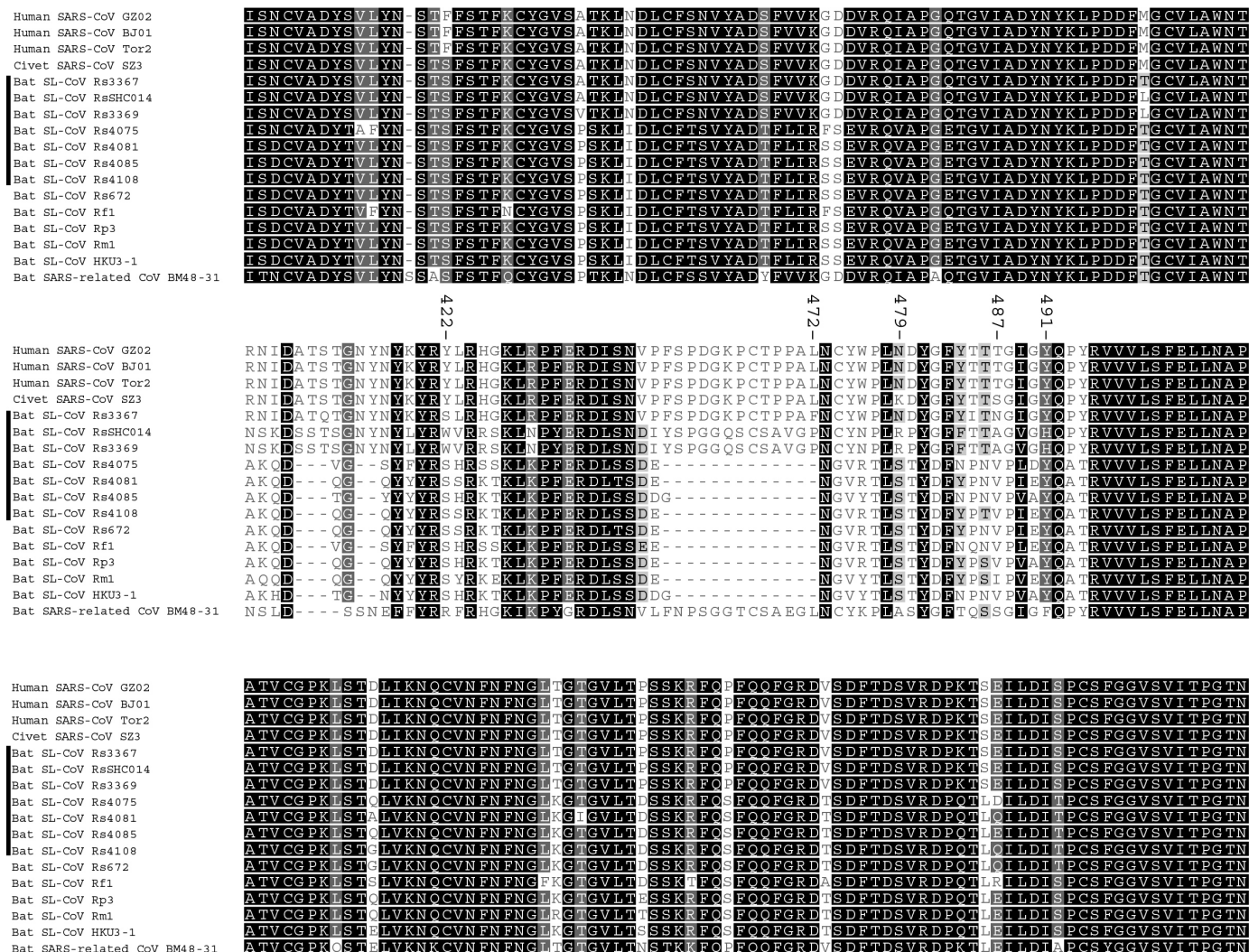
Virus infectivity detected by real-time RT–PCR. Vero E6, A549, PK15, RSKT and HeLa cells with or without expression of ACE2 of different origins were inoculated with 0.1 TCID$_{50}$ WIV-1 and incubated for 1 h at 37 °C. After removing the inoculum, the cells were cultured with medium containing 1% FBS. Supernatants were collected at 0, 12, 24 and 48 h. RNA from 140 µl of each supernatant was extracted with the Viral RNA Mini Kit (Qiagen) following manufacturer's instructions and eluted in 60 µl buffer AVE (Qiagen). RNA was quantified on the ABI StepOne system, with the TaqMan AgPath-ID One-Step RT–PCR Kit (Applied Biosystems) in a 25 µl reaction mix containing 4 µl RNA, 1 × RT–PCR enzyme mix, 1 × RT–PCR buffer, 40 pmol forward primer (5′-GTGGTGGTGACGGCA AAATG-3′), 40 pmol reverse primer (5′-AAGTGAAGCTTCTGGGCCAG-3′) and 12 pmol probe (5′-FAM-AAAGAGCTCAGCCCCAGATG-BHQ1-3′). Amplification parameters were 10 min at 50 °C, 10 min at 95 °C and 50 cycles of 15 s at 95 °C and 20 s at 60 °C. RNA dilutions from purified WIV-1 stock were used as a standard.

Serum neutralization test. SARS patient sera were inactivated at 56 °C for 30 min and then used for virus neutralization testing. Sera were diluted starting with 1:10

and then serially twofold diluted in 96-well cell plates to 1:40. Each 100 μl serum dilution was mixed with 100 μl viral supernatant containing 100 $TCID_{50}$ of WIV1 and incubated at 37 °C for 1 h. The mixture was added in triplicate wells of 96-well cell plates with plated monolayers of Vero E6 cells and further incubated at 37 °C for 2 days. Serum from a healthy blood donor was used as a negative control in each experiment. CPE was observed using an inverted microscope 2 days after inoculation. The neutralizing antibody titre was read as the highest dilution of serum which completely suppressed CPE in infected wells. The neutralization test was repeated twice.

Recombination analysis. Full-length genomic sequences of SL-CoV Rs3367 or RsSHC014 were aligned with those of selected SARS-CoVs and bat SL-CoVs using Clustal X. The aligned sequences were preliminarily scanned for recombination events using Recombination Detection Program (RDP) 4.0 (ref. 19). The potential recombination events suggested by RDP owing to their strong P values ($<10-20$) were investigated further by similarity plot and bootscan analyses implemented in Simplot 3.5.1. Phylogenetic origin of the major and minor parental regions of Rs3367 or RsSHC014 were constructed from the concatenated sequences of the essential ORFs of the major and minor parental regions of selected SARS-CoV and SL-CoVs. Two genome regions between three estimated breakpoints (20,827–26,553 and 26,554–28,685) were aligned independently using ClustalX and generated two alignments of 5,727 base pairs and 2,133 base pairs. The two alignments were used to construct maximum likelihood trees to better infer the fragment parents. All nucleotide numberings in this study are based on Rs3367 genome position.

Block 1

```
Human SARS-CoV GZ02        ISNCVADYSVLYN-STFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNT
Human SARS-CoV BJ01        ISNCVADYSVLYN-STFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNT
Human SARS-CoV Tor2        ISNCVADYSVLYN-STFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNT
Civet SARS-CoV SZ3         ISNCVADYSVLYN-STSFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNT
Bat SL-CoV Rs3367          ISNCVADYSVLYN-STSFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFTGCVLAWNT
Bat SL-CoV RsSHC014        ISNCVADYSVLYN-STSFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFLGCVLAWNT
Bat SL-CoV Rs3369          ISNCVADYSVLYN-STSFSTFKCYGVSVTKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFLGCVLAWNT
Bat SL-CoV Rs4075          ISNCVADYTAFYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRFSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rs4081          ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rs4085          ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rs4108          ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rs672           ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rf1             ISDCVADYTVFYN-STSFSTFNCYGVSPSKLIDLCFTSVYADTFLIRFSEVRQVAPGQTGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rp3             ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV Rm1             ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SL-CoV HKU3-1          ISDCVADYTVLYN-STSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRSSEVRQVAPGETGVIADYNYKLPDDFTGCVIAWNT
Bat SARS-related CoV BM48-31 ITNCVADYSVLYNSSASFSTFQCYGVSPTKLNDLCFSSVYADYFVVKGDDVRQIAPAQTGVIADYNYKLPDDFTGCVIAWNT
```

Numbered positions: 422, 472, 479, 487, 491

Block 2

```
Human SARS-CoV GZ02        RNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAP
Human SARS-CoV BJ01        RNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAP
Human SARS-CoV Tor2        RNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAP
Civet SARS-CoV SZ3         RNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLKDYGFYTTSGIGYQPYRVVVLSFELLNAP
Bat SL-CoV Rs3367          RNIDATQTGNYNYKYRSLRHGKLRPFERDISNVPFSPDGKPCTPPARNCYWPLNDYGFYITNGIGYQPYRVVVLSFELLNAP
Bat SL-CoV RsSHC014        NSKDSSTSGNYNYLYRWVRRSKLNPYERDLSNDIYSPGGQSCSAVGPNCYNPLRPYGFFTTAGVGHQPYRVVVLSFELLNAP
Bat SL-CoV Rs3369          NSKDSSTSGNYNYLYRWVRRSKINPYERDLSNDIYSPGGQSCSAVGPNCYNPLRPYGFFTTAGVGHQPYRVVVLSFELLNAP
Bat SL-CoV Rs4075          AKQD---VG--SYFYRSHRSSKLKPFERDLSSDE------------NGVRTLSTYDFNPNVPLDYQATRVVVLSFELLNAP
Bat SL-CoV Rs4081          AKQD---QG--QYYYRSSRKTKLKPFERDLTSDE------------NGVRTLSTYDFYPNVPIEYQATRVVVLSFELLNAP
Bat SL-CoV Rs4085          AKQD---TG--YYYYRSHRKTKLKPFERDLSSDDG-----------NGVYTLSTYDFNPNVPVAYQATRVVVLSFELLNAP
Bat SL-CoV Rs4108          AKQD---QG--QYYYRSSRKTKLKPFERDLSSDE------------NGVRTLSTYDFYPTVPIEYQATRVVVLSFELLNAP
Bat SL-CoV Rs672           AKQD---QG--QYYYRSSRKTKLKPFERDLTSDE------------NGVRTLSTYDFYPNVPIEYQATRVVVLSFELLNAP
Bat SL-CoV Rf1             AKQD---VG--SYFYRSHRSSKLKPFERDLSSEE------------NGVRTLSTYDFNQNVPLEYQATRVVVLSFELLNAP
Bat SL-CoV Rp3             AKQD---QG--QYYYRSHRKTKLKPFERDLSSDE------------NGVRTLSTYDFYPSVPVAYQATRVVVLSFELLNAP
Bat SL-CoV Rm1             AQQD---QG--QYYYRSYRKEKLKPFERDLSSDE------------NGVRTLSTYDFYPSIPVEYQATRVVVLSFELLNAP
Bat SL-CoV HKU3-1          AKHD---TG--NYYYRSHRKTKLKPFERDLSSDDG-----------NGVYTLSTYDFNPNVPVAYQATRVVVLSFELLNAP
Bat SARS-related CoV BM48-31 NSLD----SSNEFFYRRFRHGKIKPYGRDLSNVLFNPSGGTCSAEGLNCYKPLASYGFTQSSGIGFQPYRVVVLSFELLNAP
```

Block 3

```
Human SARS-CoV GZ02        ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Human SARS-CoV BJ01        ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Human SARS-CoV Tor2        ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Civet SARS-CoV SZ3         ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Bat SL-CoV Rs3367          ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Bat SL-CoV RsSHC014        ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Bat SL-CoV Rs3369          ATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN
Bat SL-CoV Rs4075          ATVCGPKLSTQLVKNQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGRDTSDFTDSVRDPQTLDILDITPCSFGGVSVITPGTN
Bat SL-CoV Rs4081          ATVCGPKLSTALVKNQCVNFNFNGLKGIGVLTDSSKRFQSFQQFGRDTSDFTDSVRDPQTLQILDITPCSFGGVSVITPGTN
Bat SL-CoV Rs4085          ATVCGPKLSTQLVKNQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGRDTSDFTDSVRDPQTLEILDITPCSFGGVSVITPGTN
Bat SL-CoV Rs4108          ATVCGPKLSTQLVKNQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGRDTSDFTDSVRDPQTLQILDITPCSFGGVSVITPGTN
Bat SL-CoV Rs672           ATVCGPKLSTSLVKNQCVNFNFNGFKGTGVLTDSSKRFQSFQQFGRDTSDFTDSVRDPQTLQILDITPCSFGGVSVITPGTN
Bat SL-CoV Rf1             ATVCGPKLSTSLVKNQCVNFNFNGFKGTGVLTDSSKTFQSFQQFGRDASDFTDSVRDPQTLRILDISPCSFGGVSVITPGTN
Bat SL-CoV Rp3             ATVCGPKLSTQLVKNQCVNFNFNGLKGTGVLTESSKRFQSFQQFGRDTSDFTDSVRDPQTLEILDISPCSFGGVSVITPGTN
Bat SL-CoV Rm1             ATVCGPKLSTQLVKNQCVNFNFNGIRGTGVLTSSSKRFQSFQQFGRDTSDFTDSVRDPQTLEILDISPCSFGGVSVITPGTN
Bat SL-CoV HKU3-1          ATVCGPKLSTELVKNQCVNFNFNGKGTGVLTSSSKRFQSFQQFGRDTSDFTDSVRDPQTLEILDISPCSFGGVSVITPGTN
Bat SARS-related CoV BM48-31 ATVCGPKQSTELVKNKCVNFNFNGLTGTGVLTNSTKKFQPFQQFGRDVSDFTDSVRDPKTLEILDIAPCSYGGVSVITPGTN
```

Extended Data Figure 1 | Sequence alignment of CoV S protein RBD.
SARS-CoV S protein (amino acids 310–520) is aligned with homologous regions of bat SL-CoVs using ClustalW. The newly discovered bat SL-CoVs are indicated with a bold vertical line on the left. The key amino acid residues involved in the interaction with human ACE2 are numbered on the top of the aligned sequences.

Extended Data Figure 2 | Alignment of CoV S protein S1 sequences.
Alignment of S1 sequences (amino acids 1–660) of the two novel bat SL-CoV S proteins with those of previously reported bat SL-CoVs and human and civet SARS-CoVs. The newly discovered bat SL-CoVs are boxed in red. SARS-CoV GZ02, BJ01 and Tor2 were isolated from patients in the early, middle and late phase, respectively, of the SARS outbreak in 2003. SARS-CoV

SZ3 was identified from *P. larvata* in 2003 collected in Guangdong, China. SL-CoV Rp3, Rs 672 and HKU3-1 were identified from *R. sinicus* collected in Guangxi, Guizhou and Hong Kong, China, respectively. Rf1 and Rm1 were identified from *R. ferrumequinum* and *R. macrotis*, respectively, collected in Hubei Province, China. Bat SARS-related CoV BM48-31 was identified from *R. blasii* collected in Bulgaria.
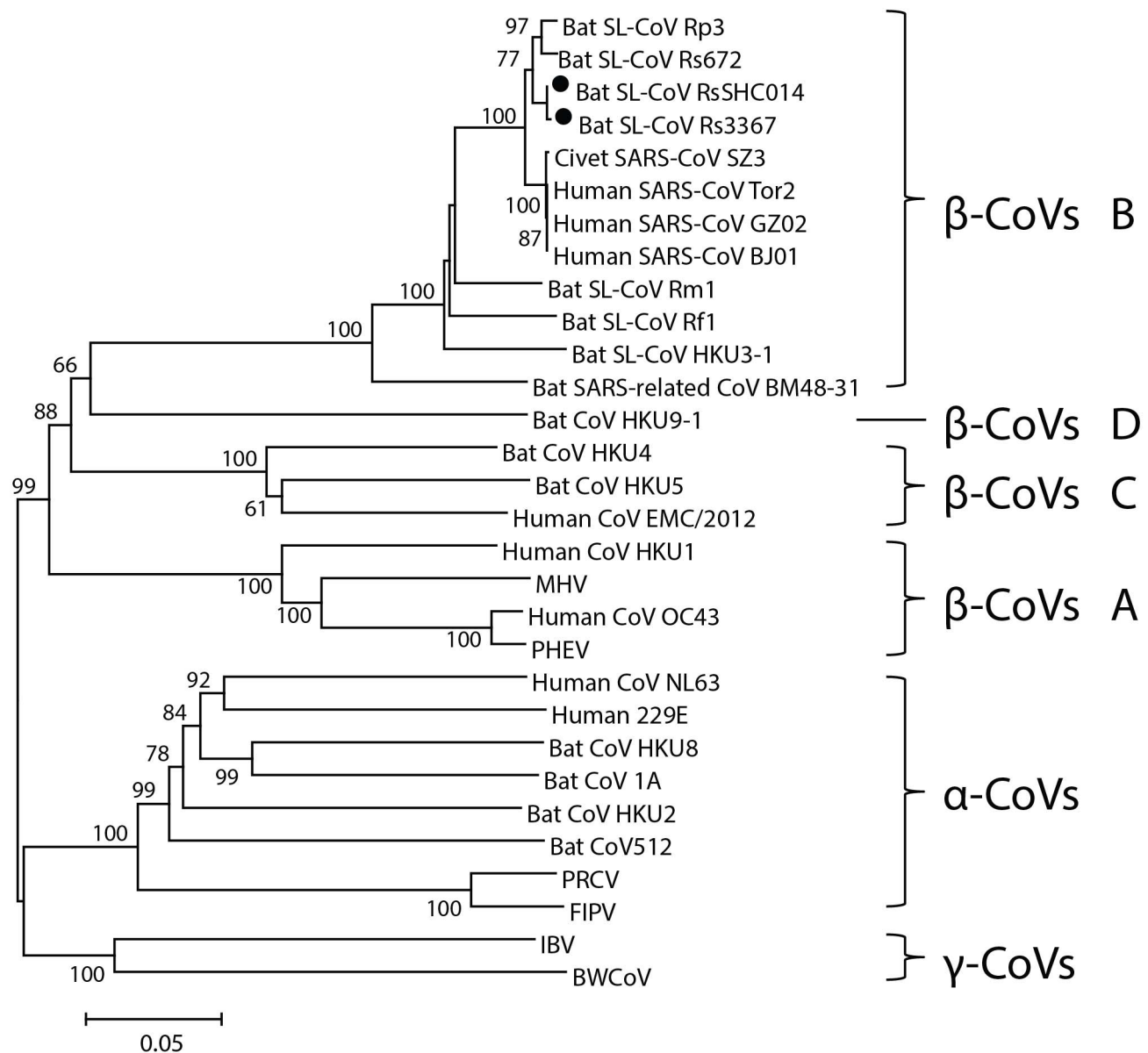
Extended Data Figure 3 | Complete RdRP sequence phylogeny.
Phylogenetic tree of bat SL-CoVs and SARS-CoVs on the basis of complete RdRP sequences (2,796 nucleotides). Bat SL-CoVs RsSHC014 and Rs3367 are highlighted by filled circles. Three established coronaivirus genera, *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus* are marked as α, β and γ, respectively. Four CoV groups in the genus *Betacoronavirus* are indicated as A, B, C and D, respectively. MHV, murine hepatitis virus; PHEV, porcine haemagglutinating encephalomyelitis virus; PRCV, porcine respiratory coronavirus; FIPV, feline infectious peritonitis virus; IBV, infectious bronchitis coronavirus; BW, beluga whale coronavirus.
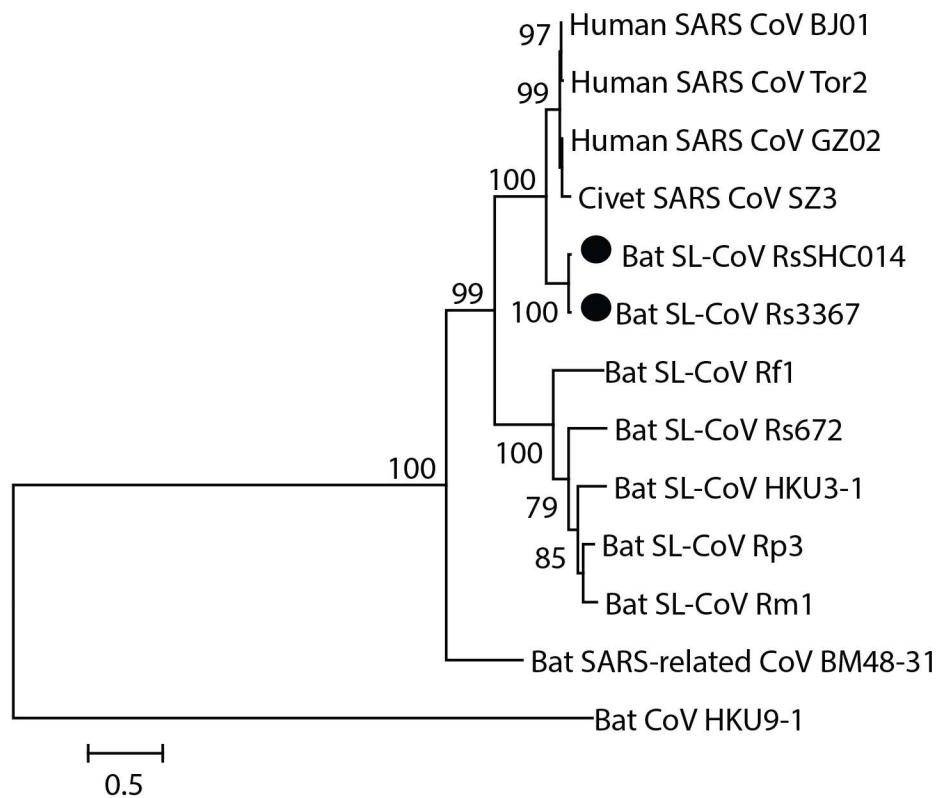
Extended Data Figure 4 | Sequence phylogeny of the complete S protein of SL-CoVs and SARS-CoV. Phylogenetic tree of bat SL-CoVs and SARS-CoVs on the basis of complete S protein sequences (1,256 amino acids). Bat SL-CoVs RsSHC014 and Rs3367 are highlighted by filled circles. Bat CoV HKU9 was used as an outgroup.
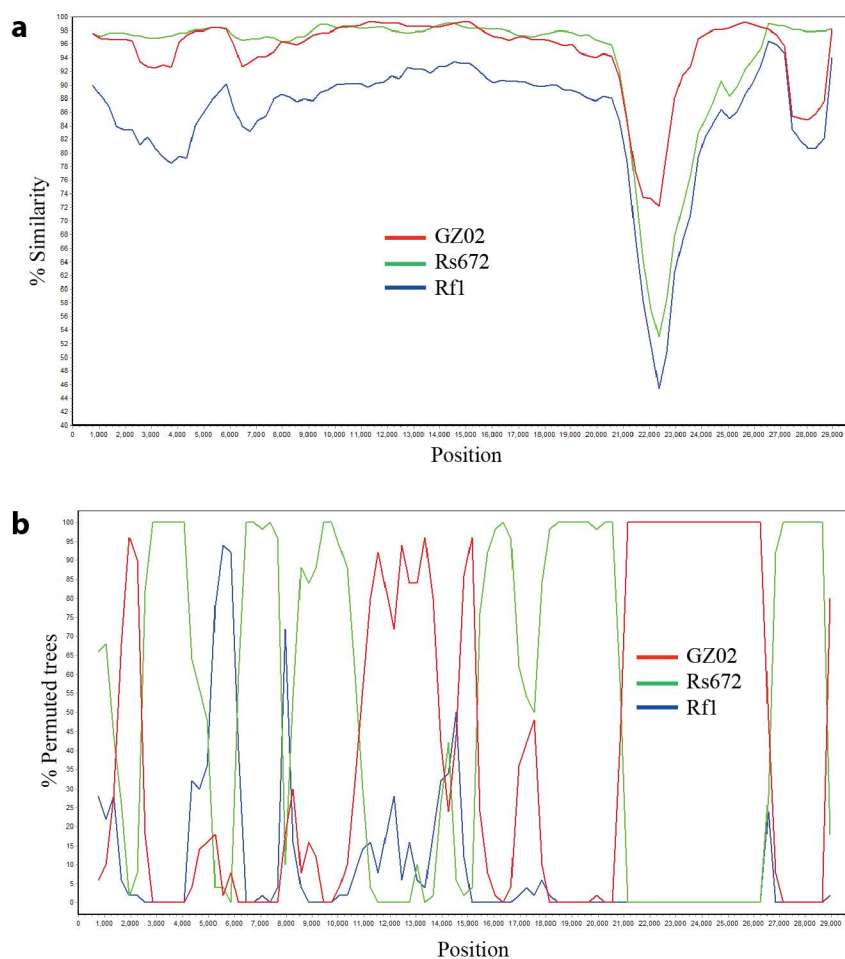
Extended Data Figure 5 | Detection of potential recombination events.
a, **b**, Similarity plot (**a**) and bootscan analysis (**b**) detected three recombination breakpoints in the bat SL-CoV Rs3367 or SHC014 genome. The three breakpoints were located at the ORF1b (nt 20,827), M (nucleotides 26,553) and N (nucleotides 28,685) genes, respectively. Both analyses were performed with an F84 distance model, a window size of 1,500 base pairs and a step size of 300 base pairs.

Extended Data Table 1 | Summary of sampling detail and CoV prevalence

Sampling time	Total number of swab or fecal samples collected	Number of CoV PCR positive samples (%)
April, 2011	14	1 (7.1)
October, 2011	10	3 (30)
May, 2012	54	4 (7.4)
September, 2012	39	19 (48.7)

Extended Data Table 2 | Genomic sequence identities of bat SL-CoVs with SARS-CoVs

		Pairwise genomic nucleotide acids identity (%)										
		Bat SARS-Like CoVs							Human and civet SARS-CoVs			
CoVs	**Genome size (nt)**	**SHC014**	**Rs672**	**Rp3**	**Rf1**	**Rm1**	**HKU3-1**	**BM48-31**	**GZ02**	**BJ01**	**Tor2**	**SZ3**
3367	29,787	98.8	92.5	93.2	87.3	88.0	87.8	76.9	95.4	95.3	95.4	95.3
SHC014	29,787	-	92.6	93.2	87.3	88.1	87.8	77.0	95.2	95.1	95.1	95.1
Rs672	29,059		-	92.4	86.2	87.4	87.0	75.2	90.9	90.9	90.8	91.0
Rp3	29,736			-	88.3	90.3	89.6	77.0	92.1	92.0	92.1	92.0
Rf1	29,709				-	89.4	88.4	76.6	87.2	87.1	87.2	87.1
Rm1	29,749					-	90.1	76.4	87.6	87.5	87.5	87.5
HKU3-1	29,728						-	76.8	87.4	87.3	87.4	87.3
BM48-31	29,276							-	76.9	77.1	77.0	76.9
GZ02	29,760								-	99.6	99.6	99.7
BJ01	29,725									-	99.8	99.6
Tor2	29,751										-	99.5
SZ3	29,741											-

Extended Data Table 3 | Genomic annotation and comparison of bat SL-CoV Rs3367 with human/civet SARS-CoVs and other bat SL-CoVs

| | | | | | ORF Identity nt/aa (%) | | | | | | | | | |
| | | | | | Human and civet SARS-CoVs | | | | Bat SARS-like CoVs | | | | | |
ORFs	Start-End (nt.)	No. of Nt.	No. of Aa.	TRS	GZ02	BJ01	Tor2	SZ3	Rs672	Rp3	Rf1	Rm1	HKU3-1	BM48-31
P1a	265-13,398	13,134	4,377	$ACGAAC_{192}AUG$	96.7/97.9	96.6/97.9	96.8/97.9	96.8/98.1	93.3/94.2	95.5/96.9	88.1/94.0	87.9/93.3	87.9/94.2	76.3/80.8
P1b	13,398-21,485	8,088	2,695		96.3/99.2	96.3/99.2	96.3/99.2	96.3/99.2	97.2/99.2	97.2/99.2	90.6/98.4	91.0/98.7	90.7/98.5	83.4/93.7
S	21,492-25,262	3,771	1,256	ACGAACAUG	88.3/90.1	88.2/90.0	88.1/89.8	88.2/90.0	76.5/78.2	76.0/79.1	74.0/77.4	76.3/79.1	75.6/78.2	70.2/74.5
(S1) *	21,493-23,535	2,043	681		78.2/81.1	78.2/80.9	78.1/80.6	78.2/81.1	65.1/62.2	63.9/63.0	62.9/62.5	64.7/63.3	65.2/63.4	62.2/64.7
(S2) *	23,536-25,263	1,728	575		98.1/99.3	98.1/99.3	98.1/99.3	98.1/99.1	87.9/94.8	88.1/95.4	85.1/92.7	87.9/95.4	86/93.5	76.6/88.2
ORF3a	25,271-26,095	825	274	$ACGAAC_{2}AUG$	99.2/98.1	98.6/97.0	98.7/97.0	98.5/96.7	90.4/90.8	84.1/84.3	88.8/86.8	83.6/84.3	83.1/82.4	72.1/71.2
ORF3b	25,692-26,036	345	114		99.1/99.1	98.2/98.2	98.2/98.2	97.9/97.3	99.1/98.2	N/D	82.6/92.1	N/D	N/D	N/D
E	26,120-26,350	231	76	$ACGAAC_{2}AUG$	98.7/98.6	98.7/98.6	98.7/98.6	98.7/98.6	99.1/98.6	97.8/98.6	96.5/96.0	96.1/97.3	97.4/98.6	91.3/93.4
M	26,401-27,066	666	221	$ACGAAC_{44}AUG$	97.4/98.1	97.2/98.1	97.2/98.1	97.2/97.7	98.7/99.5	93.3/98.1	96.3/98.6	93.2/95.4	93.9/96.8	78.5/88.1
ORF6	27,077-27,268	192	63	$ACGAAC_{155}AUG$	97.3/95.2	96.8/93.6	97.3/95.2	97.3/95.2	97.3/96.8	95.8/92.0	94.2/92.0	95.3/92	94.7/90.4	63.5/49.2
ORF7a	27,276-27,644	369	122	ACGAACAUG	94.5/95.9	94.5/95.9	94.5/95.9	94.5/95.9	97.8/100	96.2/99.1	92.9/95.0	93.4/97.5	93.2/97.5	62.3/58.1
ORF7b	27,641-27,776	135	44		96.2/93.1	96.2/93.1	96.2/93.1	96.2/93.1	99.2/100	99.2/100	97.7/97.7	99.2/100	93.3/95.4	62.9/63.6
ORF8	27,782-28,147	366	121	ACGAACAUG	47.1/46.3	N/A	N/A	47.1/46.3	97.8/100	85.2/90.2	46.2/39.0	85.7/90.2	85.7/85.3	N/A
N	28,162-29,430	1,269	422	$ACGAAC_{8}AUG$	98.3/99.5	98.4/99.5	98.4/99.5	98.4/99.5	98/98.5	96.6/97.6	93.7/95.2	96.2/97.1	95.9/96.2	77.9/87.2
s2m	29,628-29,668	41			97.5	97.5	97.5	97.5	100	100	100	100	100	95.1

*S1, the N-terminal domain of the coronavirus S protein responsible for receptor binding. S2, the S protein C-terminal domain responsible for membrane fusion.

The ORFs in the genome were predicted and potential protein sequences were translated. The pairwise comparisons were conducted for all ORFs at nucleotide acids (nt) and amino acids (aa) levels. The s2m were compared at nt level. TRS: Transcription regulating-sequences; N/D, not done; N/A, not available.

Extended Data Table 4 | Genomic annotation and comparison of bat SL-CoV RsSHC014 with human/civet SARS-CoVs and other bat SL-CoVs

ORFs	Start-End (nt.)	No. of Nt.	No. of Aa.	TRS	ORF Identity nt/aa (%)									
					Human and civet SARS-CoVs				Bat SARS-like CoVs					
					GZ02	BJ01	Tor2	SZ3	Rs672	Rp3	Rf1	Rm1	HKU3-1	BM48-31
P1a	265-13,398	13,134	4,377	ACGAAC$_{192}$AUG	96.7/97.9	96.6/97.9	96.8/97.9	96.8/98.1	93.3/94.2	95.5/96.9	88.1/94.0	87.9/93.3	87.9/94.2	76.3/80.8
P1b	13,398-21,485	8,088	2,695		96.3/99.2	96.3/99.2	96.3/99.2	96.3/99.2	97.2/99.2	97.2/99.2	90.6/98.4	91.0/98.7	90.7/98.5	83.4/93.7
S	21,492-25,262	3,771	1,256	ACGAACAUG	88.3/90.1	88.2/90.0	88.1/89.8	88.2/90.0	76.5/78.2	76.0/79.1	74.0/77.4	76.3/79.1	75.6/78.2	70.2/74.5
(S1) *	21,493-23,535	2,043	681		78.2/81.1	78.2/80.9	78.1/80.6	78.2/81.1	65.1/62.2	63.9/63.0	62.9/62.5	64.7/63.3	65.2/63.4	62.2/64.7
(S2) *	23,536-25,263	1,728	575		98.1/99.3	98.1/99.3	98.1/99.3	98.1/99.1	87.9/94.8	88.1/95.8	85.1/92.7	87.9/95.4	86/93.5	76.6/88.2
ORF3a	25,271-26,095	825	274	ACGAAC$_2$AUG	99.2/98.1	98.6/97.0	98.7/97.0	98.5/96.7	90.4/90.8	84.1/84.3	88.8/86.8	83.6/84.3	83.1/82.4	72.1/71.2
ORF3b	25,692-26,036	345	114		99.1/99.1	98.2/98.2	98.2/98.2	97.9/97.3	99.1/98.2	N/D	82.6/92.1	N/D	N/D	N/D
E	26,120-26,350	231	76	ACGAAC$_2$AUG	98.7/98.6	98.7/98.6	98.7/98.6	98.7/98.6	99.1/98.6	97.8/98.6	96.5/96.0	96.1/97.3	97.4/98.6	91.3/93.4
M	26,401-27,066	666	221	ACGAAC$_{44}$AUG	97.4/98.1	97.2/98.1	97.2/98.1	97.2/97.7	98.7/99.5	93.3/98.1	96.3/98.6	93.2/95.4	93.9/96.8	78.5/88.1
ORF6	27,077-27,268	192	63	ACGAAC$_{155}$AUG	97.3/95.2	96.8/93.6	97.3/95.2	97.3/95.2	97.3/96.8	95.8/92.0	94.2/92.0	95.3/92	94.7/90.4	63.5/49.2
ORF7a	27,276-27,644	369	122	ACGAACAUG	94.5/95.9	94.5/95.9	94.5/95.9	94.5/95.9	97.8/100	96.2/99.1	92.9/95.0	93.4/97.5	93.2/97.5	62.3/58.1
ORF7b	27,641-27,776	135	44		96.2/93.1	96.2/93.1	96.2/93.1	96.2/93.1	99.2/100	99.2/100	97.7/97.7	99.2/100	93.3/95.4	62.9/63.6
ORF8	27,782-28,147	366	121	ACGAACAUG	47.1/46.3	N/A	N/A	47.1/46.3	97.8/100	85.2/90.2	46.2/39.0	85.7/90.2	85.7/85.3	N/A
N	28,162-29,430	1,269	422	ACGAAC$_8$AUG	98.3/99.5	98.4/99.5	98.4/99.5	98.4/99.5	98/98.5	96.6/97.6	93.7/95.2	96.2/97.1	95.9/96.2	77.9/87.2
s2m	29,628-29,668	41			97.5	97.5	97.5	97.5	100	100	100	100	100	95.1

*S1, the N-terminal domain of the coronavirus S protein responsible for receptor binding. S2, the S protein C-terminal domain responsible for membrane fusion.

The ORFs in the genome were predicted and potential protein sequences were translated. The pairwise comparisons were conducted for all ORFs at nucleotide acids (nt) and amino acids (aa) levels. The s2m were compared at nt level. TRS: Transcription regulating-sequences; N/D, not done; N/A, not available.

Extended Data Table 5 | Cell lines used for virus isolation and susceptibility tests

Cell lines	Species (organ) origin	Medium	Infectivity
293T	Human (kidney)	DMEM+10%FBS	-
Hela	Human (cervix)		-
VeroE6	Monkey (kidney)		+
PK15	Pig (kidney)		+
BHK21	Hamster (kidney)		-
A549	Human (alveolar basal epithelial)		+
BK	*Myotis davidii* (kidney)	RPMI1640+10%FBS	-
RSKT	*Rhinolophus sinicus* (kidney)	DMEM/F12+10%FBS	+
MCKT	*Myotis chinensis* (kidney)		-
PaKi	*Pteropus alecto* (kidney)		-
RLK	*Rousettus leschenaulti* (kidney)		-

* Infectivity was determined by the presence of viral antigen detected by immunofluorescence assay.

LETTER

Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit

Yaser Hashem[1,2], Amedee des Georges[1,2], Vidya Dhote[3], Robert Langlois[2], Hstau Y. Liao[2], Robert A. Grassucci[2,4], Tatyana V. Pestova[3], Christopher U. T. Hellen[3] & Joachim Frank[1,2,4]

Hepatitis C virus (HCV) and classical swine fever virus (CSFV) messenger RNAs contain related (HCV-like) internal ribosome entry sites (IRESs) that promote 5′-end independent initiation of translation, requiring only a subset of the eukaryotic initiation factors (eIFs) needed for canonical initiation on cellular mRNAs[1]. Initiation on HCV-like IRESs relies on their specific interaction with the 40S subunit[2–8], which places the initiation codon into the P site, where it directly base-pairs with eIF2-bound initiator methionyl transfer RNA to form a 48S initiation complex. However, all HCV-like IRESs also specifically interact with eIF3 (refs 2, 5–7, 9–12), but the role of this interaction in IRES-mediated initiation has remained unknown. During canonical initiation, eIF3 binds to the 40S subunit as a component of the 43S pre-initiation complex, and comparison of the ribosomal positions of eIF3[13] and the HCV IRES[8] revealed that they overlap, so that their rearrangement would be required for formation of ribosomal complexes containing both components[13]. Here we present a cryo-electron microscopy reconstruction of a 40S ribosomal complex containing eIF3 and the CSFV IRES. Remarkably, although the position and interactions of the CSFV IRES with the 40S subunit in this complex are similar to those of the HCV IRES in the 40S–IRES binary complex[8], eIF3 is completely displaced from its ribosomal position in the 43S complex, and instead interacts through its ribosome-binding surface exclusively with the apical region of domain III of the IRES. Our results suggest a role for the specific interaction of HCV-like IRESs with eIF3 in preventing ribosomal association of eIF3, which could serve two purposes: relieving the competition between the IRES and eIF3 for a common binding site on the 40S subunit, and reducing formation of 43S complexes, thereby favouring translation of viral mRNAs.

Canonical translation initiation begins with assembly of a 43S pre-initiation complex, comprising a 40S subunit, eIF1, eIF1A, the initiator methionine transfer RNA (Met-tRNA$_i^{Met}$)–eIF2–GTP ternary complex (eIF2-TC) and the approximately 800-kDa five-lobed multi-subunit eIF3 (ref. 1). The 43S complex attaches to the cap-proximal region of mRNA and then scans to the initiation codon, whereupon it forms a 48S initiation complex with established codon-anticodon base-pairing. Attachment and scanning are mediated by eIF4A, eIF4B and eIF4F, but scanning on structured mRNAs additionally requires DHX29 (refs 14, 15), a DExH-box protein that also binds directly to the 40S subunit[13,14]. 48S complex formation on the homologous HCV and CSFV IRESs, which comprise two principal domains, II and III (Extended Data Fig. 1a), does not involve scanning and requires only a 40S subunit and the eIF2-TC. The process is based on the specific interaction of the IRES with the 40S subunit, which involves the IRES pseudoknot and subdomains IIId and IIIe[3–5,16]. Binding to the 40S subunit positions the initiation codon of the IRES in the P site, where it directly base-pairs with the anticodon of Met-tRNA$_i^{Met}$ as a part of the eIF2-TC, leading to formation of the 48S complex. Subsequent joining of the 60S subunit to this complex is mediated by eIF5 and eIF5B. Although domain II of

HCV-like IRESs stimulates eIF5-mediated hydrolysis of eIF2-bound GTP and joining of a 60S subunit[17–19], it does not influence the affinity of the IRES for the 40S subunit[5], only moderately affects 48S complex formation, and is not essential for initiation on the CSFV IRES[3,18,20,21].

An unresolved aspect of initiation on HCV-like IRESs is the role of eIF3, which interacts specifically with the apical region of domain III
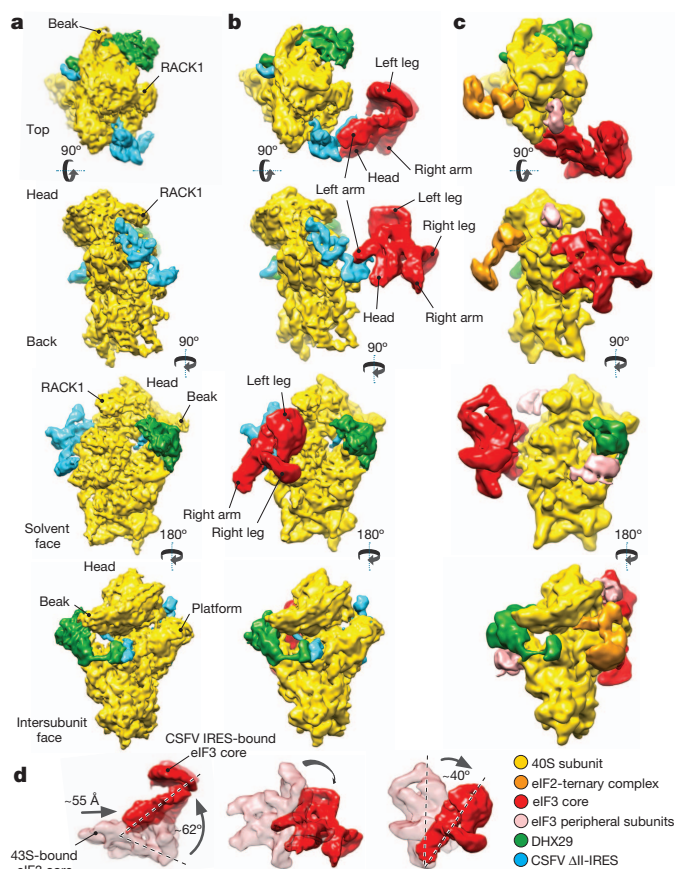
Figure 1 | Cryo-electron microscopy structures of the CSFV ΔII-IRES–40S–DHX29 complex alone and bound to eIF3 compared to the structure of the DHX29-bound 43S preinitiation complex. a, CSFV ΔII-IRES–40S–DHX29 complex (class 2, Extended Data Fig. 3). **b**, CSFV ΔII-IRES–40S–DHX29 complex bound to eIF3 (class 4, Extended Data Fig. 3). **c**, 43S preinitiation complex[13]. Complexes are viewed from the top, the back, the solvent and the intersubunit faces from top to bottom, respectively. In panels **a–c**, the 40S subunit is displayed in yellow, DHX29 in green, the eIF3 structural core in red and the CSFV ΔII-IRES in cyan. **d**, Comparison between different positions and orientations of eIF3 in the 43S complex and in the CSFV ΔII-IRES–40S complexes, as indicated.

[1]Howard Hughes Medical Institute (HHMI), Department of Biochemistry and Molecular Biophysics, Columbia University, New York City, New York 10032, USA. [2]Department of Biochemistry and Molecular Biophysics, Columbia University, New York City, New York 10032, USA. [3]Department of Cell Biology, SUNY Downstate Medical Center, Brooklyn, New York 11203, USA. [4]Department of Biological Sciences, Columbia University, New York City, New York 10032, USA.
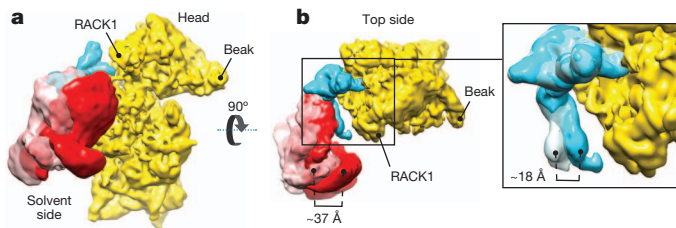
Figure 2 | Different orientations of eIF3 and subdomain IIIb in the CSFV ΔII IRES–40S–DHX29 complex. a, Solvent-side view of eIF3 in the two most divergent orientations, as it appears in classes 4 (solid red surface) and 6 (transparent pink surface) of the CSFV ΔII-IRES–40S–eIF3 complex, bound to the CSFV ΔII-IRES (cyan) on the 40S subunit (yellow). **b,** Left, top view of eIF3 in the two most divergent orientations, bound to the CSFV ΔII-IRES on the 40S subunit. **b,** Right, blow-up focused on domain IIIb of the CSFV IRES, showing the extent of its reorientation. The brackets display the magnitude of the movement of eIF3 and of IRES domain IIIb in the two most divergent orientations.

(helices IIIb and III₄)[5,9,11] (Extended Data Fig. 1a). Although eIF3 is not essential for 48S complex formation on these IRESs and only slightly stimulates this process in the *in vitro* reconstituted translation system[2,6,7,18], mutations in the apical region of domain III that impair binding of eIF3 (refs 2, 9, 19) lead to severe translation initiation defects in cell-free extracts[22]. Importantly, the position of the eIF3 core in 43S complexes[13] and of the HCV IRES in 40S–IRES binary complexes[8] overlap, with a clash between the left arm of eIF3 and the pseudoknot[13]. The simultaneous presence of eIF3 and the IRES in ribosomal complexes would therefore require their rearrangement.

To shed light on the role of eIF3 in initiation on HCV-like IRESs and to investigate how the predicted eIF3/IRES clash is resolved, we determined the cryo-electron microscopy structure of the 40S subunit in complex with eIF3 and the CSFV IRES lacking the non-essential domain II (ΔII-IRES). The CSFV IRES was chosen because it has higher translational activity than the HCV IRES[2], probably because it interacts more strongly with eIF3 and/or the 40S subunit, and would thus yield complexes with higher stability for structural analysis. Domain II was omitted to reduce complexity and to reduce conformational heterogeneity. DHX29 was also included in these complexes because it stabilizes the peripheral domains of eIF3 in 43S complexes[13] without affecting the interaction of the IRES with the 40S subunit[14]. The 40S–ΔII-IRES–eIF3–DHX29 complexes were assembled *in vitro* by incubating individual purified components. Toeprinting analysis of these complexes revealed that they maintained the full complement of interactions of the IRES with the 40S subunit and eIF3 (refs 2, 18), and were quantitatively converted into 48S complexes upon addition of eIF2-TC (Extended Data Fig. 2). Although DHX29 does not interfere with 48S complex formation on the ΔII-IRES (Extended Data Fig. 2), only 10% of 43S complexes in cells are bound to DHX29 (ref. 14), and whether it is present in IRES-bound ribosomal initiation complexes in the cytoplasm has not been determined. Processing of approximately 630,000 particle images (see Methods) yielded several classes containing different combinations of components (Extended Data Figs 3 and 4). The present analysis is focused on the 40S–DHX29–ΔII-IRES complex (class 2, ~72,900 particles) and the first of three 40S–DHX29–ΔII-IRES–eIF3 classes (4 to 6), which differed slightly in the orientation of eIF3 and eIF3-bound subdomain IIIb of the IRES (class 4, ~26,000 particles). They yielded 8.5 Å and 9.3 Å reconstructions, respectively, which revealed three well-defined densities on the 40S subunit (Fig. 1a, b).

The shape and location of a density around the tip of h16 and of a smaller mass on the subunit interface near the A site connected to it via a clearly defined linker (green in Fig. 1a, b) matched the density of DHX29 in the 43S complex[13] (Fig. 1c). Another density at the back of the platform (cyan in Fig. 1a, b) was assigned to the CSFV ΔII-IRES, because it fitted the shape and location of the related HCV IRES domain III in 40S–IRES and 80S–IRES binary complexes[8,23] (Extended

Data Fig. 1b). A visible segment of the CSFV coding region emerges from the mRNA entrance and forms a small additional mass near the A site and the small intersubunit domain of DHX29. This mass could be modelled as an approximately 11-base-pairs stem-loop separated from the P site codon by 5 to 7 nucleotides and could include a predicted hairpin formed by CSFV nucleotides 387–406.

The third mass attached to the ΔII-IRES at the back of the 40S subunit (red in Fig. 1b) was attributed to the core of eIF3, on the basis of its shape[12,13,24] and interaction with the apical part of IRES domain III[2,5,9,11]. Remarkably, the position of the eIF3 core in the IRES-containing complex differs from that in the 43S complex. Whereas in the 43S complex the left arm and head of eIF3 interact with ribosomal proteins rpS1e/rpS26e and rpS13e/rpS27e, respectively[13] (Fig. 1c), in the IRES-containing complex they bind to the apical part of IRES domain III (Fig. 1b), consistent with the position of eIF3 on HCV-like IRESs[5,9,11]. The orientation of the CSFV ΔII-IRES on eIF3 is consistent with the position of scattered density attributed to the HCV IRES in a lower-resolution cryo-electron microscopy reconstruction of the eIF3–IIIabc binary complex[11]. Thus, the IRES effectively usurps ribosomal contacts of eIF3, leading to displacement of eIF3 from the 40S subunit and leaving it interacting exclusively with the IRES. Compared to its position in the 43S complex, the eIF3 core is shifted by approximately 55 Å and rotated by approximately 60° (Fig. 1d). Assignment of the left arm and head of eIF3 to eIF3a and eIF3c subunits, respectively[24] (Extended Data Fig. 5a, b), makes interaction of the IRES with the left arm and head of eIF3 consistent both with reports of ultraviolet cross-linking of the HCV IRES to eIF3a and eIF3c subunits[9–11] and with the observed impairment of binding of eIF3 to the HCV IRES by mutations in eIF3a and eIF3c[11]. Interestingly, eIF3 and the subdomain IIIb bound to it adopt a number of orientations in the 40S–DHX29–ΔII-IRES–eIF3 complex (classes 4 to 6). Between the two most divergent orientations, shown by classes 4 and 6, the tip of domain IIIb moves by up to 18 Å, inducing in turn a movement of the eIF3 core that reaches up to 37 Å in the region of the legs (Fig. 2a, b and Supplementary Video 1). Thus, the inherent flexibility of the IIIabc–III₄ four-way junction[25] allows the eIF3-bound domain IIIb to move while the IRES maintains its contacts with the 40S subunit. In the class 4 map, the high variance and the low resolution in the region of eIF3 legs, when compared to the rest of the eIF3 core (Extended Data Fig. 5c), points to a continuum of orientations of subdomain IIIb and eIF3. The same explanation probably
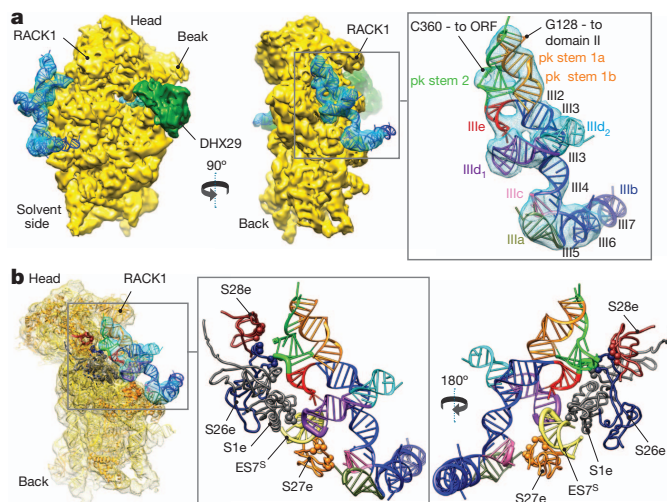
Figure 3 | Structure and atomic model of the CSFV ΔII-IRES bound to the 40S subunit. a, Atomic model of the CSFV ΔII-IRES fitted into its density map (blue mesh), seen from the solvent (left) and back (middle) sides. Right panel displays a blow-up on the CSFV ΔII-IRES atomic model (ribbon), coloured variably to highlight its different subdomains. **b,** Ribosomal proteins contacting the CSFV ΔII-IRES when bound to the 40S subunit, seen from the back (left and middle) and the front (right).

applies to the other two classes, 5 and 6, on the basis of the lower resolution observed in the eIF3 legs. Despite the presence of DHX29, none of the peripheral domains of eIF3 was clearly observed, which is probably a consequence of the displacement of the eIF3 core from its binding site on the 40S subunit.

The sub-nanometre resolution of the class 2 map allowed us to attempt pseudo-atomic modelling of the CSFV ΔII-IRES (nucleotides 129–361), for which no high-resolution structure is available, to our knowledge. The model (Fig. 3a) was built and fitted into its density mass segmented from both class 2 and class 4 reconstructions (bound with eIF3 in position 1), yielding a final cross-correlation coefficient with the classes 2 and 4 IRES densities of 0.92 and 0.93, respectively (see Methods). The model is consistent with the results of phylogenetic comparisons, chemical/enzymatic probing and mutational analyses of CSFV and related IRESs[3,20,26]. The apical region consists of a long cylindrical stem, from which the subdomains IIId1 and IIId2 protrude, and is kinked at the flexible elbow formed by the four-way junction of helix III$_4$ and subdomains IIIa, IIIb and IIIc. Subdomains IIIb and IIId2 extend away from the 40S subunit surface, consistent with their lack of involvement in 40S subunit binding[2,3,19]. Domain IIIb is not well resolved in the 40S–DHX29–ΔII-IRES complex but is stabilized in the complex containing eIF3. The basal region of domain III is formed by the pseudoknot, subdomain IIIe and helix III$_1$, which together form two sets of coaxially stacked helices, angled at ∼40° with respect to each other, that are directly comparable to the 'main' and 'sidecar' helices of the analogous region in the HCV IRES[27] (Extended Data Fig. 6).

The higher resolution of the present maps also allows confident assignment of individual interactions of the IRES with the 40S subunit and eIF3. In addition to the mRNA flanking the initiation codon, five distinct elements of the ΔII-IRES (subdomains IIIa, IIIc, IIId1, IIIe and the pseudoknot) contact the 40S subunit (Fig. 3b and Extended Data Fig. 7a). These elements are highly conserved in HCV-like IRESs in members of several genera of Flaviviridae and Picornaviridae (for example ref. 28), and each of them coincides precisely with sites in HCV and CSFV IRESs that are protected from enzymatic cleavage or chemical modification by the bound 40S subunit[3–5,16]. On the 40S subunit, the interactions mostly involve ribosomal proteins (rpS1e, rpS26e, rpS27e and rpS28e) (Fig. 3b and Extended Data Fig. 7a) and, consistent with these contact sites, interactions of corresponding elements of the HCV IRES with rpS1e and rpS27e have been observed[29,30]. Importantly, rpS1e, rpS26e and rpS27e are also involved in the interaction of the 40S subunit with eIF3 (ref. 13), which accounts for the displacement of eIF3 from it by the IRES. However, the apical loop of subdomain IIId1 also contacts the apical loop of ES7 of 18S ribosomal RNA (Fig. 3b), probably through base-pairing between the conserved GGG nucleotides on subdomain IIId1 and CCC nucleotides in the apical loop of ES7. This interaction, which induces a small-scale shift in the position of the apex of ES7 towards the head, was previously suggested by a low-resolution cryo-electron microscopy study[23], and is consistent with the strong ribosomal protection of this region of the IRES in footprinting studies[3–5,16]. Interestingly, the ribosomal elements that HCV-like IRESs exploit for binding to the 40S subunit (rpS1e, rpS26e, rpS27e, rpS28e and ES7) are all eukaryote-specific.

The conserved GGG motif in subdomain IIId1 is a major determinant of ribosome binding and initiation activity for all HCV-like IRESs[3,5,6]. However, whereas the apical UCCC loop of ES7 is highly conserved in vertebrates, the equivalent element in plants has the sequence CUUA, which would not base-pair stably to the GGG motif, probably contributing to the inability of wheat 40S subunits to bind to the HCV IRES[2]. The observed interactions with the 40S subunit also account for the severe effects of substitutions in the apical loop of subdomain IIId[3,5,6,16,19] and of disruption of base-pairing in pseudoknot stem 2 (refs 2, 5, 20, 26).
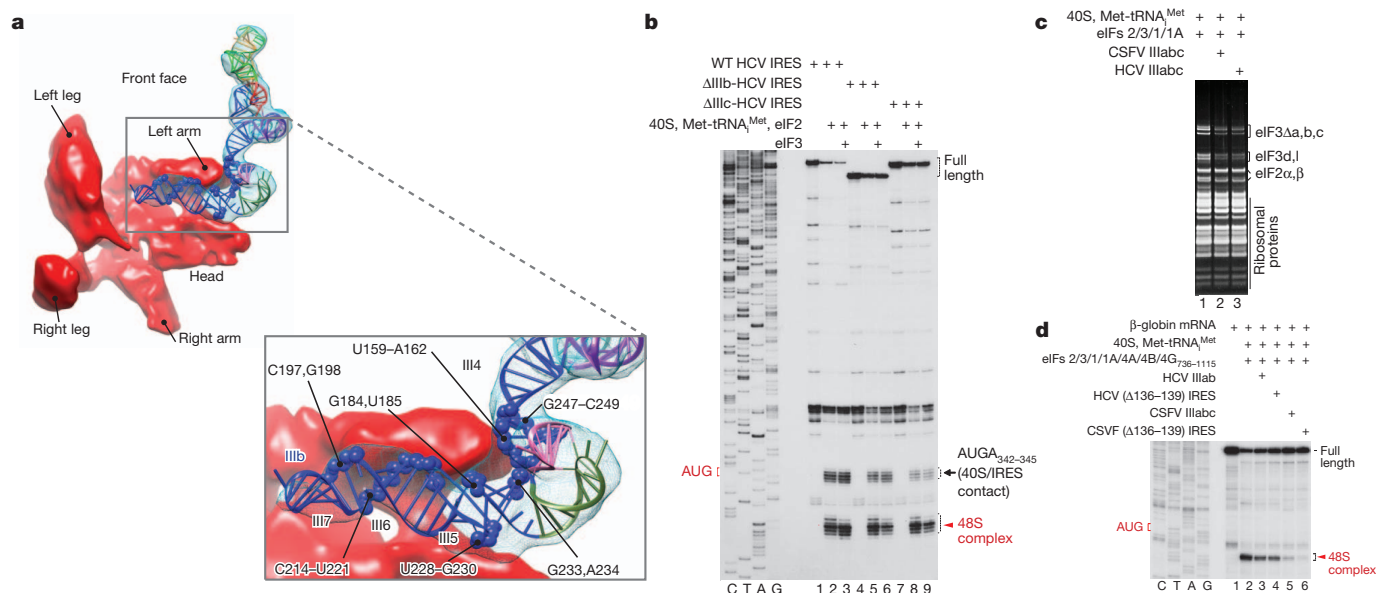
Figure 4 | Binding of eIF3 to subdomain IIIb of the CSFV IRES and effects on translation of the eIF3–IRES interaction. a, eIF3 binding site on the CSFV IRES. The residues potentially interacting with eIF3 from domain IIIb in the CSFV IRES are highlighted by blue spheres and labelled (bottom). **b,** Toeprinting analysis of 48S complexes assembled on wild-type (WT) HCV (nucleotides 1–349)–CAT mRNA (WT HCV IRES) and ΔIIIb-HCV IRES or ΔIIIc-HCV IRES derivatives[22] lacking either domain IIIb (nucleotides 172–227) or domain IIIc (nucleotides 229–238) with 40S subunits, Met-tRNA$_i^{Met}$, eIF2 and eIF3 as indicated. Primer extension was arrested at nucleotides 342–345 by stably bound 40S subunits[2] and at nucleotides 355–359 by 48S complexes, as indicated. Lanes C, T, A and G show the cDNA sequence corresponding to WT HCV IRES mRNA. The position of the initiation codon AUG$_{373}$ is indicated on the left. **c,** Inhibition of 43S complex formation by subdomain IIIabc of HCV and CSFV IRESs, assayed by sucrose density gradient centrifugation (SDG). The protein composition of ribosomal peak fractions was analysed by SDS–PAGE and fluorescent SYPRO staining. **d,** Inhibition of 48S complex formation on β-globin mRNA by IIIabc subdomains and by complete HCV and CSFV IRESs containing a 4-nucleotide deletion in helix III$_2$, assayed by toeprinting. Lanes C, T, A and G show the complementary DNA sequence corresponding to β-globin mRNA. The position of the initiation codon is indicated on the left. Each gel reported in the figure is representative of results obtained from three technical replicates.

Binding of the CSFV ΔII-IRES to the eIF3 core is mainly restricted to two regions, helix III$_4$ and domain IIIb (Fig. 4a). Residues G247–C249 and U159–A162 in helix III$_4$, residues G184–U185 and G233–A234 in helix III$_5$, residues C214–U221 in helix III$_6$ and residues C197–G198 in helix III$_7$ are all in close proximity to eIF3's left arm (eIF3a) and likely interact with it directly. Residues U228–G230 seem to contact eIF3's head (eIF3c) and are thus also likely involved in binding of the IRES to eIF3. These interactions are consistent with the position of eIF3 on HCV and CSFV IRESs determined by footprinting[5,9,11] (Extended Data Fig. 7b–d). Our finding that the interaction of eIF3 with the CSFV IRES primarily involves the left arm of eIF3 (eIF3a) is supported by the observations that mutations in eIF3a have a stronger effect than mutations in eIF3c on the binding of eIF3 to the HCV IRES[11].

The fact that in CSFV ΔII-IRES–eIF3-containing ribosomal complexes, the 40S subunit interacts only with the IRES, which in turn also binds to eIF3 through its ribosome-binding surface, suggests that in the case of the HCV-like IRES mutants lacking the eIF3-binding site, eIF3 would more readily compete with them for the conventional site on the 40S subunit, thus reducing 48S complex formation. To test this prediction, we compared 48S complex formation in the presence and in the absence of eIF3 on the wild-type HCV IRES and those IRES mutants that lacked subdomains IIIb or IIIc and could therefore no longer bind eIF3 (refs 2, 22). Consistent with the prediction of our hypothesis, inclusion of eIF3 reduced 48S complex formation on the mutants, whereas in the presence of only 40S subunits and the eIF2-TC, the level of 48S complex formation on all three mRNAs was very similar (Fig. 4b). The small stimulatory effect of eIF3 on 48S complex formation on the wild-type HCV IRES could be due to stabilization of the IRES structure at the junction of domains IIIa, b and c, which in turn might stabilize interaction between IIIa and rpS27e. This stabilization could have a more significant function in the cell by counteracting the dissociative influence of eIF1 on 48S complexes formed on HCV-like IRESs[18]. However, eIF3 is not essential for subsequent stages in initiation, because 48S complexes formed in its absence on both wild-type and ΔIIIb mutant HCV IRESs readily underwent subunit joining, forming elongation-competent 80S ribosomes (Extended Data Fig. 8).

Consistent with the IIIabc domain of HCV-like IRESs and the 40S subunit binding to a common site on eIF3, HCV and CSFV IIIabc domains impaired 43S complex formation by reducing ribosomal association of eIF3 by 60–70% (Fig. 4c). Consequently, inclusion of these domains in reaction mixtures strongly inhibited 48S complex formation on β-globin mRNA (Fig. 4d). Inhibition by IIIabc domains was almost as potent as by complete IRESs with a 4-nucleotide deletion in helix III$_2$, which could no longer bind 40S subunits but retained eIF3-binding activity[2]. The CSFV IIIabc domain was a stronger inhibitor, paralleling the higher translational activity of CSFV IRES in cell-free extracts[2], which could therefore be due to its greater ability to compete for eIF3.

In conclusion, our unexpected finding that the CSFV ΔII-IRES displaces the core of eIF3 from its position on the 40S subunit sheds light on the role of eIF3's interaction with HCV-like IRESs in the mechanism of initiation, and provides a plausible explanation for why mutations in the apical region of domain III that impair binding of eIF3 lead to severe translation defects in cell-free extracts[22]. Thus, by binding to eIF3, HCV-like IRESs would reduce the competition with this factor for binding to the 40S subunit and would also impair formation of 43S complexes, which in turn might aid the ability of these IRESs to compete with cellular mRNAs.

METHODS SUMMARY

For cryo-electron microscopy studies, 40S–DHX29–IRES–eIF3 complexes were assembled *in vitro* using CSFV ΔII-IRES mRNA[3], native eIF2, eIF3 and 40S subunits purified from rabbit reticulocyte lysate, and recombinant DHX29. Single-particle cryo-electron microscopy studies were done as described[13], with further details given in Supplementary Information. 48S complex formation on wild-type and mutant HCV IRESs[22] and on β-globin mRNA was assayed by toeprinting using native eIF2, eIF3 and 40S subunits and recombinant eIF1, eIF1A, eIF4A, eIF4B and

eIF4G$_{736–1115}$ as described[2,14]. The influence of subdomain IIIabc of HCV and CSFV IRESs on 43S complex formation was assayed by sucrose density gradient centrifugation[14].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Rev. Mol. Cell Biol.* **11,** 113–127 (2010).
2. Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev.* **12,** 67–83 (1998).
3. Kolupaeva, V. G., Pestova, T. V. & Hellen, C. U. Ribosomal binding to the internal ribosomal entry site of classical swine fever virus. *RNA* **6,** 1791–1807 (2000).
4. Kolupaeva, V. G., Pestova, T. V. & Hellen, C. U. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *J. Virol.* **74,** 6242–6250 (2000).
5. Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7,** 194–206 (2001).
6. de Breyne, S., Yu, Y., Pestova, T. V. & Hellen, C. U. Factor requirements for translation initiation on the Simian picornavirus internal ribosomal entry site. *RNA* **14,** 367–380 (2008).
7. Pisarev, A. V. *et al.* Functional and structural similarities between the internal ribosome entry sites of hepatitis C virus and porcine teschovirus, a picornavirus. *J. Virol.* **78,** 4487–4497 (2004).
8. Spahn, C. M. *et al.* Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* **291,** 1959–1962 (2001).
9. Sizova, D. V., Kolupaeva, V. G., Pestova, T. V., Shatsky, I. N. & Hellen, C. U. Specific interaction of eukaryotic translation initiation factor 3 with the 5′ nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J. Virol.* **72,** 4775–4782 (1998).
10. Buratti, E., Tisminetzky, S., Zotti, M. & Baralle, F. E. Functional analysis of the interaction between HCV 5′UTR and putative subunits of eukaryotic translation initiation factor eIF3. *Nucleic Acids Res.* **26,** 3179–3187 (1998).
11. Sun, C. *et al.* Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation. *Nucleic Acids Res.* **41,** 7512–7521 (2013).
12. Siridechadilok, B., Fraser, C. S., Hall, R. J., Doudna, J. A. & Nogales, E. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* **310,** 1513–1515 (2005).
13. Hashem, Y. *et al.* Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell* **153,** 1108–1119 (2013).
14. Pisareva, V. P., Pisarev, A. V., Komar, A. A., Hellen, C. U. & Pestova, T. V. Translation initiation on mammalian mRNAs with structured 5′UTRs requires DExH-box protein DHX29. *Cell* **135,** 1237–1250 (2008).
15. Abaeva, I. S., Marintchev, A., Pisareva, V. P., Hellen, C. U. & Pestova, T. V. Bypassing of stems versus linear base-by-base inspection of mammalian mRNAs during ribosomal scanning. *EMBO J.* **30,** 115–129 (2011).
16. Lukavsky, P. J., Otto, G. A., Lancaster, A. M., Sarnow, P. & Puglisi, J. D. Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nature Struct. Biol.* **7,** 1105–1110 (2000).
17. Locker, N., Easton, L. E. & Lukavsky, P. J. HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J.* **26,** 795–805 (2007).
18. Pestova, T. V., de Breyne, S., Pisarev, A. V., Abaeva, I. S. & Hellen, C. U. eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J.* **27,** 1060–1072 (2008).
19. Ji, H., Fraser, C. S., Yu, Y., Leary, J. & Doudna, J. A. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proc. Natl Acad. Sci. USA* **101,** 16990–16995 (2004).
20. Fletcher, S. P. & Jackson, R. J. Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5′ untranslated region important for IRES function. *J. Virol.* **76,** 5024–5033 (2002).
21. Friis, M. B., Rasmussen, T. B. & Belsham, G. J. Modulation of translation initiation efficiency in classical swine fever virus. *J. Virol.* **86,** 8681–8692 (2012).
22. Rijnbrand, R. *et al.* Almost the entire 5′ non-translated region of hepatitis C virus is required for cap-independent translation. *FEBS Lett.* **365,** 115–119 (1995).
23. Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* **13,** 1695–1706 (2005).
24. Querol-Audi, J. *et al.* Architecture of human translation initiation factor 3. *Structure* **21,** 920–928 (2013).
25. Melcher, S. E., Wilson, T. J. & Lilley, D. M. The dynamic nature of the four-way junction of the hepatitis C virus IRES. *RNA* **9,** 809–820 (2003).
26. Rijnbrand, R., van der Straaten, T., van Rijn, P. A., Spaan, W. J. & Bredenbeek, P. J. Internal entry of ribosomes is directed by the 5′ noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J. Virol.* **71,** 451–457 (1997).
27. Berry, K. E., Waghray, S., Mortimer, S. A., Bai, Y. & Doudna, J. A. Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. *Structure* **19,** 1456–1466 (2011).

28. Hellen, C. U. & de Breyne, S. A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse picornavirus genera: evidence for modular exchange of functional noncoding RNA elements by recombination. *J. Virol.* **81,** 5850–5863 (2007).

29. Babaylova, E. *et al.* Positioning of subdomain IIId and apical loop of domain II of the hepatitis C IRES on the human 40S ribosome. *Nucleic Acids Res.* **37,** 1141–1151 (2009).

30. Malygin, A. A., Shatsky, I. N. & Karpova, G. G. Proteins of the human 40S ribosomal subunit involved in hepatitis C IRES binding as revealed from fluorescent labeling. *Biochemistry (Moscow)* **78,** 53–59 (2013).

Supplementary Information is available in the online version of the paper.

Author Contributions Y.H., A.d.G., V.D, T.V.P., C.U.T.H. and J.F. interpreted the data and wrote the manuscript. V.D. and T.V.P. prepared the sample. Y.H., A.d.G. and R.A.G. performed the cryo-electron microscopy experiments. H.Y.L. performed the three-dimensional variance estimation. Y.H. and R.L. performed the cryo-electron microscopy data processing. Y.H. modelled the CSFV IRES. T.V.P., C.U.T.H. and J.F. directed research.

Author Information The electron microscopy map has been deposited in the EMBL-European Bioinformatics Institute Electron Microscopy Data Bank under accession codes EMD-2450 and EMD-2451. Coordinates of electron microscopy-based model of the CSFV IRES have been deposited in the RCSB Protein Data Bank under accession number 4c4q. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.U.T.H (christopher.hellen@downstate.edu) and J.F. (jf2192@columbia.edu).

METHODS

Plasmids. The plasmid HCV (MSTN-STOP) was made (GenScript) by inserting a 497-nucleotide DNA fragment between XbaI and EcoRI sites of pUC57 that consisted of a T7 promoter sequence, HCV type 1b nucleotides 40–375 (modified to include a UAA stop codon in place of the fifth coding triplet and a BamHI restriction site 20 nucleotides downstream of it), followed by a segment of the influenza nonstructural (NS) protein coding sequence[2].

The plasmid ΔIIIb HCV (MSTN-STOP) was constructed similarly, except that the HCV sequence contained a deletion of nucleotides 172–227 (corresponding to subdomain IIIb) in addition to the UAA stop codon and BamHI restriction site.

The plasmid HCV(Δ136–139) was also constructed similarly, except that the wild-type HCV sequence was modified to include a deletion of nucleotides 136–139, a ScaI restriction site 2 nucleotides upstream of the HCV initiation codon and HindIII and BamHI restriction sites 13 nucleotides and 32 nucleotides downstream of it, respectively.

The plasmid HCV(IIIabc) was made (GenScript) by inserting a 142-nucleotide DNA fragment between XbaI and SalI sites of pUC57 that consisted of a T7 promoter sequence and HCV type 1b nucleotides 143–250, followed by a NaeI restriction site.

The plasmids pWT-CAT, pΔE-CAT and pΔF-CAT[22] consist of nucleotides 1–349 of the HCV-H strain, or variants thereof lacking subdomain IIIb (nucleotides 172–227) or subdomain IIIc (nucleotides 229–238) respectively, linked to a CAT reporter cistron.

The plasmid CSFV(Δ145-148) was made (GenScript) by inserting a 556 nucleotides DNA fragment between XbaI and HindIII sites of pUC57 that consisted of a T7 promoter sequence, CSFV (Alfort/Tuebingen) nucleotides 1–442 (modified to delete nucleotides 145–148, ref. 2, and to include a FspI restriction site 6 nucleotides upstream of the CSFV initiation codon and a DraI restriction site 12 nucleotides downstream of it), followed by a segment of the influenza NS coding sequence[2].

The plasmid CSFV(IIIabc) was made (GenScript) by inserting a 137-nucleotide DNA fragment between XbaI and HindIII sites of pUC57 that consisted of a T7 promoter sequence and CSFV (Alfort/Tuebingen) nucleotides 152–256, followed by a SmaI restriction site.

The plasmid pCSFV(128-442)NS′[3] was generated from pCSFV(1-442)NS′[2] and consists of CSFV (Alfort/Tuebingen) nucleotides 128–442 linked to a segment of the influenza NS coding sequence.

The MVHL-Stop[31] modified β-globin transcription vector consists of a DNA fragment corresponding to a T7 promoter sequence, four CAA repeats and the complete human β-globin sequence, modified to convert the 5th coding triplet to a UAA termination codon and the downstream UGUGU sequence to AGUGA, cloned between BglII and XhoI sites of pET28a.

CSFV ΔII-IRES mRNA was transcribed after linearization of pCSFV(128-442)NS′[3] with PmlI, 235 nucleotides downstream of the CSFV initiation codon. The wild-type, ΔIIIb (Δ172–227 nucleotides) and ΔIIIc (Δ229–238 nucleotides) HCV IRES mRNAs were transcribed from pWT-CAT, pΔE-CAT and pΔF-CAT plasmids, respectively, that had been linearized with HindIII. HCV (MSTN-STOP) and ΔIIIb HCV (MSTN-STOP) mRNAs were transcribed after linearization of corresponding plasmids with EcoRI. HCV(Δ136-139), CSFV(Δ145-148), HCV IIIabc and CSFV IIIabc mRNAs were transcribed after linearization of corresponding plasmids with ScaI, FspI, SalI and HindIII, respectively. All mRNAs were transcribed using T7 RNA polymerase.

Purification of ribosomal subunits, initiation and elongation factors, DHX29 and aminoacylation of tRNA. Native 40S and 60S ribosomal subunits, eIF2, eIF3, eIF5B, eEF1H, eEF2 and total aminoacyl-tRNA synthetases were purified from rabbit reticulocyte lysate, and recombinant human DHX29, eIF1, eIF1A, eIF4A, eIF4B and eIF4G$_{736–1115}$, and *Escherichia coli* methionyl tRNA synthetase were expressed and purified from *E. coli as* described previously[31–33]. Native total rabbit tRNA (Novagen) was aminoacylated with Met, Ser, Thr and Asn using native aminoacyl-tRNA synthetases, whereas *in-vitro*-transcribed tRNA$_i$^{Met} (ref. 34) was aminoacylated using *E. coli* methionyl tRNA synthetase as described[32].

Assembly of 40S–eIF3–DHX29 ΔII-IRES complexes for cryo-electron microscopy analysis. Complexes for cryo-electron microscopy analysis were assembled by incubating 20 pmol 40S subunits, 30 pmol eIF3, 30 pmol DHX29 and 24 pmol CSFV ΔII-IRES RNA in 50 μl of buffer containing 20 mM Tris pH 7.5, 75 mM KCl, 5 mM MgCl$_2$, 2 mM DTT and 0.25 mM spermidine for 10 min at 37 °C. Before applying onto grids, the reaction mixture was diluted with the same buffer to the concentration of 40S subunits of 32 nM.

Toeprinting analysis of 48S initiation and 80S pre-termination complex (pre-TC) formation on wild-type and mutant CSFV and HCV IRES mRNAs. 48S complexes were assembled by incubating 2 pmol mRNA with 2 pmol 40S subunits, 4 pmol eIF2 and 5 pmol Met-tRNA$_i$^{Met}, 3 pmol eIF3 and 2 pmol DHX29, as indicated, in 20 μl buffer A (20 mM Tris pH 7.5, 100 mM KCl, 2.5 mM MgCl$_2$, 2 mM DTT, 0.25 mM spermidine) supplemented with 1 mM ATP and 0.4 mM GTP for

10 min at 37 °C. For 80S initiation complex formation, reaction mixtures were supplemented with 4 pmol 60S subunits, 4.5 pmol eIF5 and 2 pmol eIF5B, and incubated at 37 °C for an additional 10 min to allow formation of 80S initiation complexes. To form pre-TCs, 80S initiation complexes were supplemented with 4 pmol eEF1H, 12 pmol eEF2 and ~10 μg appropriately aminoacylated total native tRNA, and incubated at 37 °C for an additional 10 min. Ribosomal complexes were analysed by toeprinting[32] using avian myeloblastosis virus (AMV) reverse transcriptase and a ^{32}P-labelled primer. cDNA products were resolved in 6% polyacrylamide sequencing gels.

Toeprinting analysis of 48S complex formation on β-globin mRNA. 48S complexes were assembled by incubating 2 pmol of a derivative of β-globin mRNA containing four 5′-terminal CAA repeats (MVHL-STOP mRNA[31]) with 2 pmol 40S subunits, 4 pmol eIF2, 5 pmol Met-tRNA$_i$^{Met}, 3 pmol eIF3, 5 pmol eIF4A, 2 pmol eIF4B, 3 pmol eIF4G$_{736–1115}$, 10 pmol eIF1, 10 pmol eIF1A in the presence and in the absence of 15 pmol IIIabc subdomains and by complete HCV and CSFV IRESs containing a 4-nucleotide deletion in helix III$_2$ in 20 μl buffer A supplemented with 1 mM ATP and 0.4 mM GTP for 10 min at 37 °C. Ribosomal complexes were analysed by toeprinting[32] using AMV reverse transcriptase and a ^{32}P-labelled primer. cDNA products were resolved in 6% polyacrylamide sequencing gels.

Inhibition of 43S complex formation by subdomain IIIabc of HCV and CSFV IRESs assayed by sucrose density gradient (SDG) centrifugation. 43S complexes were assembled by incubating 20 pmol 40S subunits, 50 pmol eIF2, 70 pmol Met-tRNA$_i$^{Met}, 100 pmol eIF1, 100 pmol eIF1A and 30 pmol eIF3 in the presence and in the absence of 150 pmol IIIabc subdomains of HCV and CSFV IRESs in 200 μl buffer A supplemented with 1 mM ATP and 0.4 mM GTP for 10 min at 37 °C. The reaction mixtures were then subjected to centrifugation through a 10–30% SDG prepared in buffer A in a Beckman SW55 rotor at 53,000 r.p.m. for 1 h 15 min. Fractions that corresponded to 43S ribosomal complexes were analysed by SDS–PAGE with subsequent fluorescent SYPRO (Molecular Probes) staining.

Electron microscopy. Four microlitres of each sample was applied to holey carbon grids (carbon-coated Quantifoil 2/4 grid, Quantifoil Micro Tools) bearing an additional continuous thin layer of carbon[35]. Grids were blotted and vitrified by rapidly plunging into liquid ethane at −180 °C with a Vitrobot (FEI)[36,37]. Data acquisition was done under low-dose conditions (12 e$^-$ Å^{-2}) on a FEI Tecnai F20 (FEI, Eindhoven) operating at 120 kV with a Gatan CT3500 side-entry cryo-holder. The data set was collected automatically using Leginon[38] at a calibrated magnification of ×51,570 on a 4k × 4k Gatan Ultrascan 4000 CCD camera with a physical pixel size of 15 μm, thus making the pixel size 2.245 Å on the object scale.

Image processing. The data were preprocessed using pySPIDER (R.L. and J.F., unpublished) and Arachnid. Arachnid is a Python-encapsulated version of SPIDER[39], replacing SPIDER batch files with Python. It also contains novel procedures such as Autopicker, which was used for the automated particle selection, yielding a total number of particles of ~630,000, picked from ~12,000 micrograph images. Those selected particles were classified with RELION[40], ultimately yielding six classes (see Details on the three-dimensional classification and Extended Data Fig. 9): Class 1: 40S–ΔII-IRES, ~56,000 particles; Class 2: 40S–ΔII-IRES–DHX29, ~72,900 particles; Class 3: 40S–DHX29–eIF3, ~18,000 particles; Class 4: 40S–ΔII-IRES–DHX29–eIF3, ~26,000 particles, where eIF3 is in 'orientation 1'; Class 5: 40S–ΔII-IRES–DHX29–eIF3, ~18,000 particles, where eIF3 is in 'orientation 2', and Class 6: 40S–ΔII-IRES–DHX29–eIF3, ~17,000 particles, where eIF3 is in 'orientation 3'. We focused our analysis on classes 2 and 4, scoring resolutions of 8.5 Å and 9.5 Å, respectively, estimated with the gold standard Fourier shell correlation (FSC) = 0.143 (Extended Data Fig. 4a)[40,41]. To assess the quality of our reconstruction, we performed a reference-free two-dimensional classification using RELION[40] and compared the obtained class-averages with projections generated from our final reconstruction (Extended Data Fig. 4b).

Details on the three-dimensional classification. The unsupervised three-dimensional classification of IRES-bound ribosomal complexes consisted of six rounds of three-dimensional classifications (Extended Data Fig. 9) conducted in a quasi-hierarchical fashion, using RELION[40] (version 1.2b7). The classes generated by each round were analysed and either regrouped and reclassified in a subsequent round or rejected because of their structural inconsistency with the known structures of the different components of the complex. At the end of the different rounds of classifications, particles from similar classes were regrouped and refined together as one class. For each run of classification and refinement, the small ribosomal 40S subunit[42] (PDBID: 2XZM) was used as an initial reference. The reference was generated by simulating a cryo-electron microscopy density map from the atomic coordinates file of the 40S subunit using UCSF Chimera[43]. The reference map was filtered to 40 Å for classification and auto-refinement runs. For all classification runs, the regularization factor T in RELION was set to 3.

The first run of classification had the purpose of eliminating those data windows containing obvious contaminants from the rest of the data set and was set for 10 classes, with the following sampling parameters: angular sampling interval of 30°,

an offset search range of 21 pixels and an offset search step of 3 pixels. The sampling parameters were progressively narrowed in the course of the 36 classification iterations, to 3.7° for the angular sampling interval, 8 pixels for the offset search range and 2 pixels for the offset search step. At the end of the first classification round, three classes (representing 33% of the 630,000 picked windows of particles) were inconsistent with the known structure of the 40S ribosomal subunit and were thus rejected (labelled 'rejects'). Out of the first round of classification, particles from 7 classes (representing 67% of the data set) were pooled together for the second round of classification.

The second round of three-dimensional classification and all of the subsequent rounds of classification started with the following initial sampling parameters: angular sampling interval of 15°, an offset search range of 14 pixels and an offset search step of 2 pixels. The sampling parameters were progressively narrowed in the course of the 42 classification iterations, to 1.8° for the angular sampling interval, 4 pixels for the offset search range and 1 pixel for the offset search step. All 10 classes of this second classification round were considered to be potentially consistent with the structure of the 40S subunit and thus no rejects were singled out. Based on the visual analysis of these classes, it was possible to regroup structurally similar classes of particles into two groups. Two separate three-dimensional classification rounds were conducted in parallel on the particles of each group, round 3 and round 4.

Round 3 was performed using particles collated from 7 different classes (representing 44% of the full data set) originating from the previous round, round 2 (Extended Data Fig. 9), and was set to generate 8 classes. This round of classification was conducted for 46 iterations and resulted in 4 classes of particles inconsistent with the structure of the 40S subunit and/or eIF3 and the CSFV IRES (classes 3 to 6 representing 21% of the full data set, Extended Data Fig. 9).

Round 4 of classification was performed using particles pooled from 2 different classes (representing 23% of the full data set) originating from the previous round (round 2, Extended Data Fig. 9) and was set to generate 8 classes.

Based on the visual analysis of the classes derived from rounds 3 and 4, we regrouped some of their different classes into two groups and the particles from structurally similar classes forming each group were collated. Two other separate three-dimensional classification rounds were conducted in parallel on the particles of each group, round 5 and round 6.

Round 5 was performed using particles from class 2 originating from classification round 3 (representing 5% of the full data set, Extended Data Fig. 9) and due to the low number of particles, it was set to generate 4 classes only. This round was conducted for 37 iterations.

Round 6 was performed using particles from classes 6 and 7 originating from round 4 and class 1 originating from round 3 (representing 14% of the full data set) and was set to generate 8 classes. The round yielded 7 classes of rejects (representing 13% of the full data set). Classes 1, 3, 5 and 6 were rejected because of the inconsistency of the shape of eIF3 with its known structure, which appears to be bound to the CSFV IRES but assumed a scattered and/or deformed aspects, probably due to a very large degree of flexibility. Classes 2, 4 and 7 were rejected because of the low-resolution appearance of the 40S subunit and the inconsistency of the shape of the CSFV IRES with its structure.

Based on the similarities among different classes originating from rounds 3, 4, 5 and 6, particles from certain classes were pooled and auto-refined, using RELION's Auto-Refine module, into six final classes (Extended Data Fig. 9) as follows: 1, particles from class 8 of round 3, and from class 8 of round 4 (9% of the full data set) displaying 40S–ΔII-IRES complexes; 2, particles from class 7 of round 3, and from classes 3 and 5 of round 4 (12% of the full data set) displaying 40S–ΔII-IRES–DHX29 complexes; 3, particles from class 4 of round 4 and from class 8 of round 6 (3% of the full data set) displaying 40S–DHX29–eIF3 complexes; 4, particles from class 2 of round 5 and from class 2 of round 4 (4% of the full data set) displaying 40S–ΔII-IRES–DHX29–eIF3, where eIF3 is in 'orientation 1'; 5, particles from class 3 of round 5 and from class 1 of round 4 (3% of the full data set) displaying 40S–ΔII-IRES–DHX29–eIF3, where eIF3 is in 'orientation 2'; 6, particles from classes 1 and 4 of round 5 (3% of the full data set) displaying 40S–ΔII-IRES–DHX29–eIF3, where eIF3 is in 'orientation 3'. In all of the auto-refinements, RELION was set with the following initial sampling parameters: angular sampling interval of 15°, an offset search range of 14 pixels and an offset search step of 2 pixels. The number of iterations for each refinement was determined automatically by RELION based on the improvement of the resolution between consecutive iterations.

Three-dimensional variance estimation. The particle windows were binned threefold in order to reduce the memory and CPU requirements for variance estimation, yielding a pixel size of 6.735 Å. The three-dimensional variance map was computed using the bootstrapping method[44–47] for the class 4 map (Extended Data Fig. 3) presenting eIF3 in orientation 1, as follows: forty thousand bootstrap reconstructions were generated, each of which was obtained from $N = 26,317$

particle projections that were randomly sampled with replacement from the total set of the N particles. The bootstrap volumes were filtered to about twice the first zero-crossing of the contrast transfer functions (CTFs) to boost the signal-to-noise ratio of the three-dimensional variance. The structural variance was estimated as the sample variance of the bootstrap volumes minus the variance of the noise, and the difference was then multiplied by N (ref. 44). In this estimation, the noise variance is assumed to be uniform across the map[45].

Density maps segmentation and display. Cryo-electron microscopy reconstructions were segmented using the SEGGER module[48] implemented in UCSF Chimera[43]. Segments counting less than 10,000 voxels were discarded. Segments were refined manually using the VOLUME ERASER module implemented in UCSF Chimera. Finally, the obtained segments were smoothed using a Gaussian filter in the VOLUME FILTER module also implemented in Chimera. The final maps were displayed and rendered with Chimera.

CSFV IRES modelling and fitting. The ΔII-IRES RNA was modelled based on the established secondary structure of the CSFV IRES[20]. The secondary structure was loaded into the S2S nucleic acid alignment and modelling tool[49] and the CSFV IRES secondary structure was exported to Assemble, a nucleic acid two-dimensional/three-dimensional modelling tool[50]. As domain II is absent in ΔII-IRES RNA, only domain III was modelled from the IRES sequence (nucleotides 129–361; GenBank J04358) and a three-dimensional model was generated in Assemble and placed into the electron microscopy map.

The model was relaxed and fitted into the IRES map using Molecular Dynamics Flexible Fitting (MDFF)[51]. MDFF is an MD simulation-based fitting procedure, which applies an extra potential to the system, related to the gradient of the cryo-electron microscopy density map. The initial system was prepared for MDFF using VMD[52] and consisted of the atomic model of the CSFV ΔII-IRES and its corresponding segmented map from class 2 particles, which correspond to the 40S–ΔII-IRES–DHX29 complex. As the model was built into the electron microscopy map directly, no rigid-body fitting was required. To achieve a better representation of the inter- and intra-molecular interactions, the system was embedded in a solvent box of TIP3P water molecules, with an extra 12 Å padding in each direction, and neutralized by potassium ions, and an excess of ∼ 0.2 M KCl was added. The system was minimized for 2,000 steps in NAMD[53] followed by MDFF in explicit solvent. The run was stopped after 400 ps of simulation time, when the cross-correlation coefficient between the model and the map, as well as the root mean squared deviation of the model during the trajectories had stabilized. The simulated system was prepared using CHARMM force field parameters (Combined CHARMM All-Hydrogen Topology File for CHARMM22 Proteins and CHARMM27 Lipids[54,55]). The same protocol was reapplied in order to fit the CSFV ΔII-IRES into its corresponding density segmented from class 4, 40S–ΔII-IRES–DHX29–eIF3, where eIF3 is in orientation 1. This last MDFF endeavours to fit flexible domain IIIb of the CSFV IRES into one specific conformation in interaction with eIF3 in order to identify the IRES residues interacting with the latter (Fig. 4a).

To identify the CSFV IRES binding site on the body of the 40S subunit in terms of ribosomal proteins and rRNA, the crystal structure of *Tetrahymena thermophila* 40S subunit[42] was rigid-body fitted into the class 2, 40S–ΔII-IRES–DHX29, using UCSF Chimera[43]. The highest CCC guided the optimal fit.

31. Skabkin, M. A. *et al.* Activities of ligatin and MCT-1/DENR in ribosomal translation initiation and ribosomal recycling. *Genes Dev.* **24,** 1787–1801 (2010).
32. Pisarev, A. V., Unbehaun, A., Hellen, C. U. & Pestova, T. V. Assembly and analysis of eukaryotic translation initiation complexes. *Methods Enzymol.* **430,** 147–177 (2007).
33. Lomakin, I. B., Shirokikh, N. E., Yusupov, M. M., Hellen, C. U. & Pestova, T. V. The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *EMBO J.* **25,** 196–210 (2006).
34. Pestova, T. V. & Hellen, C. U. Preparation and activity of synthetic unmodified mammalian tRNA$_i^{Met}$ in initiation of translation in vitro. *RNA* **7,** 1496–1505 (2001).
35. Grassucci, R. A., Taylor, D. J. & Frank, J. Preparation of macromolecular complexes for cryo-electron microscopy. *Nature Protocols* **2,** 3239–3246 (2007).
36. Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21,** 129–228 (1988).
37. Wagenknecht, T., Frank, J., Boublik, M., Nurse, K. & Ofengand, J. Direct localization of the tRNA-anticodon interaction site on the *Escherichia coli* 30 S ribosomal subunit by electron microscopy and computerized image averaging. *J. Mol. Biol.* **203,** 753–760 (1988).
38. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151,** 41–60 (2005).
39. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
40. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
41. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20,** 205–214 (2012).
42. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331,** 730–736 (2011).

43. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
44. Penczek, P. A., Yang, C., Frank, J. & Spahn, C. M. Estimation of variance in single-particle reconstruction using the bootstrap technique. *J. Struct. Biol.* **154,** 168–183 (2006).
45. Zhang, W., Kimmel, M., Spahn, C. M. & Penczek, P. A. Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis. *Structure* **16,** 1770–1776 (2008).
46. Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455,** 416–420 (2008).
47. Liao, H. Y. & Frank, J. Classification by bootstrapping in single particle methods. *Proc. IEEE Int. Symp. Biom. Imaging* 169–172 (2010).
48. Pintilie, G. D., Zhang, J., Goddard, T. D., Chiu, W. & Gossard, D. C. Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170,** 427–438 (2010).
49. Jossinet, F. & Westhof, E. Sequence to Structure (S2S): display, manipulate and inter-connect RNA data from sequence to structure. *Bioinformatics* **21,** 3320–3321 (2005).
50. Jossinet, F., Ludwig, T. E. & Westhof, E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26,** 2057–2059 (2010).
51. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16,** 673–683 (2008).
52. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph. Model.* **14,** 33–38 (1996).
53. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
54. Brooks, B. R. *et al.* CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4,** 187–217 (1983).
55. MacKerell, A. D. Jr *et al.* CHARMM: the energy function and its parameterization with an overview of the program. *Enc. Comput. Chem.* **1,** 271–277 (1998).
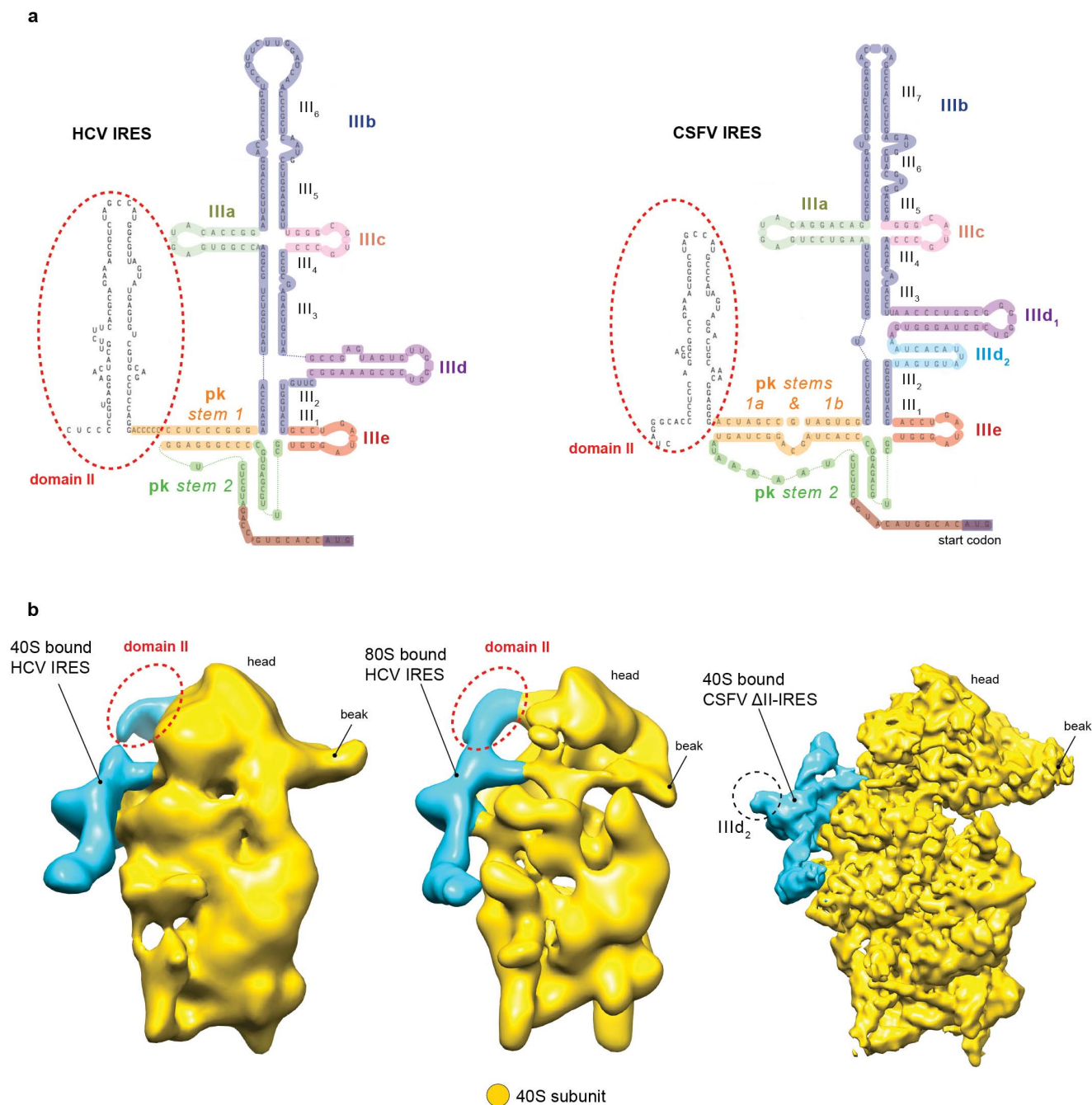
Extended Data Figure 1 | Comparison of HCV and CSFV IRES-bound ribosomal complexes. a, Secondary structures of (left) the HCV IRES and (right) the CSFV IRES. Domain II of each IRES is indicated by a red dashed oval; elements of the pseudoknot and subdomains IIIa–IIIe are colour-coded as in Extended Data Fig. 6. **b**, Cryo-electron microscopy reconstructions of the HCV IRES bound to the rabbit 40S subunit at 20 Å resolution[8] (left), the HCV IRES bound to the 40S subunit of cycloheximide-stalled human 80S ribosomes at 15 Å resolution[23] (middle) (accession code EMD-1138) and the CSFV

ΔII-IRES bound to the rabbit 40S subunits at 8.5 Å resolution (right) (this study). In all panels, the IRES-40S subunit is viewed from the solvent side; the 40S subunit is displayed in yellow and the IRES in cyan. The red dashed circles in left and middle panels show a discontinuity in the density of domain II in the HCV IRES bound to the 40S subunit compared to the HCV IRES bound to 80S ribosomes. The dashed circle in the right hand panel highlights CSFV IRES subdomain IIId$_2$, which has no counterpart in the HCV IRES.
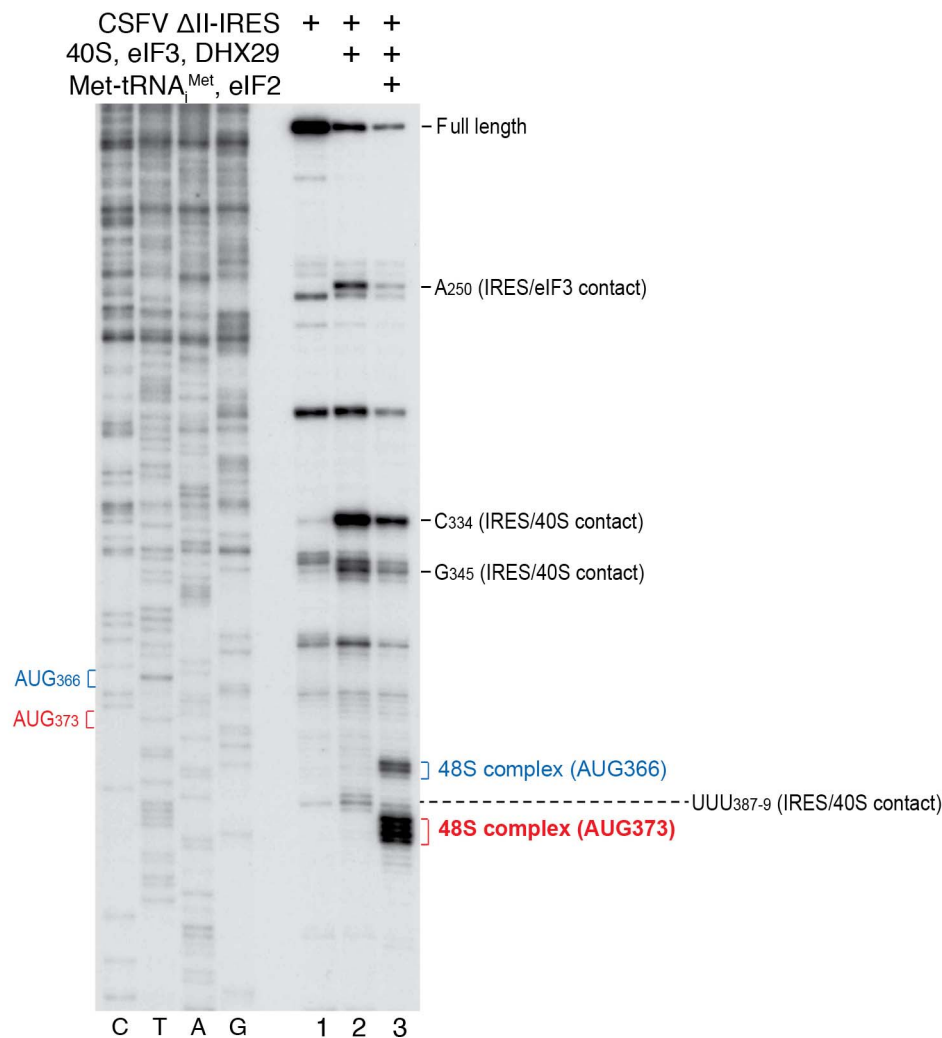
Extended Data Figure 2 | Analysis of 40S–ΔII-IRES–eIF3–DHX29 complexes. 40S–ΔII-IRES–eIF3–DHX29 complexes were assembled *in vitro* using CSFV ΔII-IRES mRNA, native eIF2, eIF3 and 40S subunits purified from rabbit reticulocyte lysate and recombinant DHX29, and assayed by toeprinting. Lanes C, T, A and G show the cDNA sequence corresponding to CSFV ΔII-IRES mRNA. The position of the initiation codon is indicated on the left. This analysis revealed (lane 2) that deletion of domain II of the IRES or the presence of DHX29 did not influence IRES's contacts with either 40S subunit (the toeprint stops at $UUU_{387-389}$, G_{345} and C_{334}) or eIF3 (the toeprint stops at A_{250}) that have been previously observed[2,18]. Moreover, upon addition of the eIF2-TC, 40S–ΔII-IRES–eIF3–DHX29 complexes were quantitatively converted into 48S complexes on the authentic initiation codon AUG_{373} (lane 3). The low-efficiency 48S complex formation on the preceding AUG_{366} was also observed before and was not related to the presence of DHX29[18]. The gel reported in the figure is representative of results obtained from three technical replicates.

class 1
40S●IRES
~56,000
particles

class 2
40S●IRES●
DHX29
~72,900
particles

class 3
40S●DHX29●
eIF3
~ 18,000
particles

~ 12,000 micrographs
~ 630,000 particles

Classification

rejects
~ 423,000 particles
inconsistent with
40S structure

class 4
40S●IRES●
DHX29●eIF3
Orientation 1
~26,000
particles

class 5
40S●IRES●
DHX29●eIF3
Orientation 2
~18,000
particles

class 6
40S●IRES●
DHX29●eIF3
Orientation 3
~17,000
particles
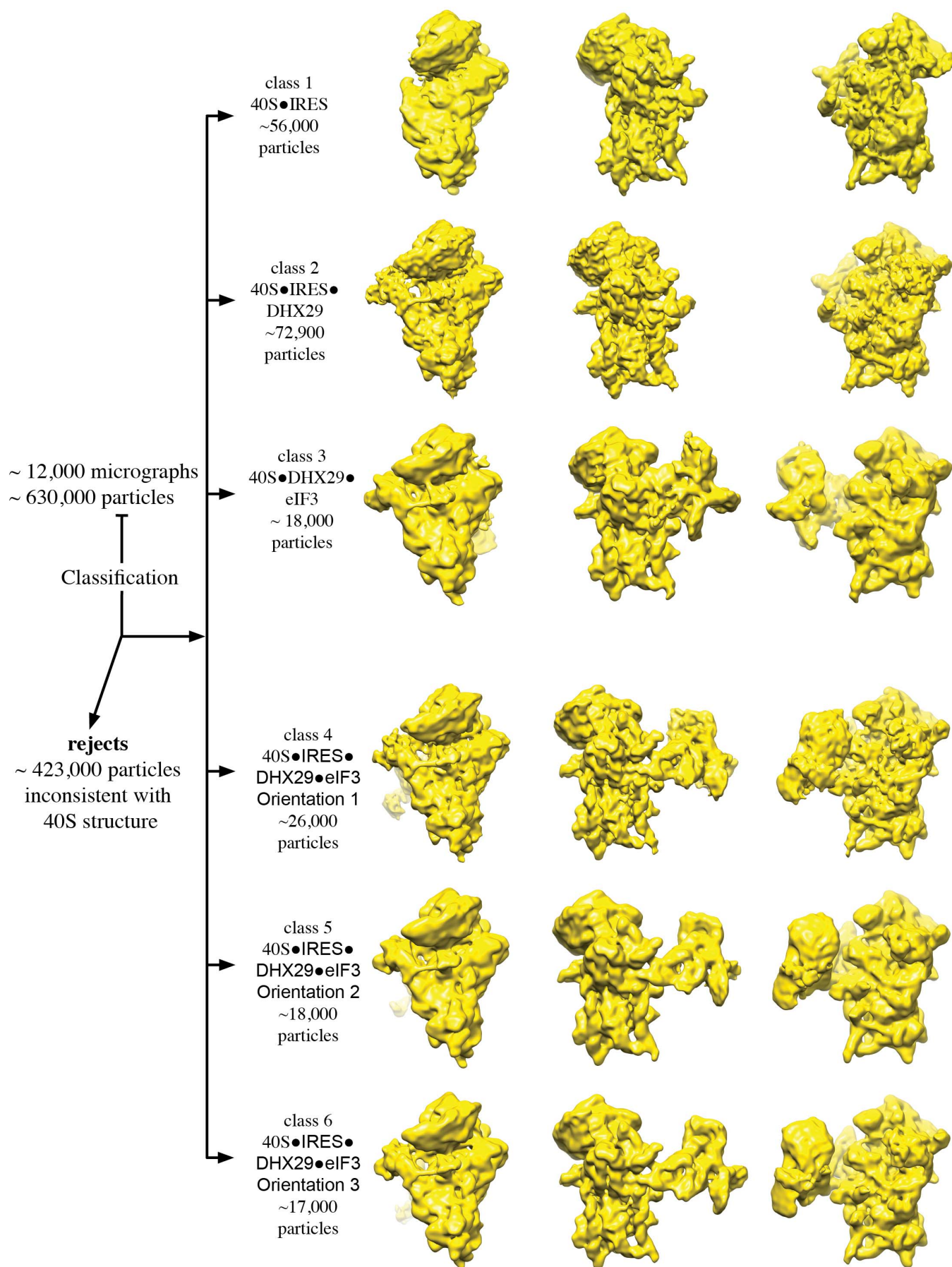
Extended Data Figure 3 | Unsupervised three-dimensional classification of IRES-bound ribosomal complexes. Unsupervised three-dimensional classification of IRES-bound ribosomal complexes identified ~423,000 particles inconsistent with the known structure of the 40S subunit (rejects) and six well-populated classes containing complexes of the 40S subunit in a binary complex with the ΔII-IRES (class 1), of the 40S subunit bound to the ΔII-IRES and DHX29 (class 2), of the 40S subunit bound to DHX29 and eIF3 (class 3) and of the 40S subunit bound to the ΔII-IRES, DHX29 and eIF3 in orientation 1 (class 4), in orientation 2 (class 5) and in orientation 3 (class 6), viewed from (left) the back, (centre) the intersubunit side and (right) the solvent side.

a

Gold-standard Fourier Shell Correlation curves

b
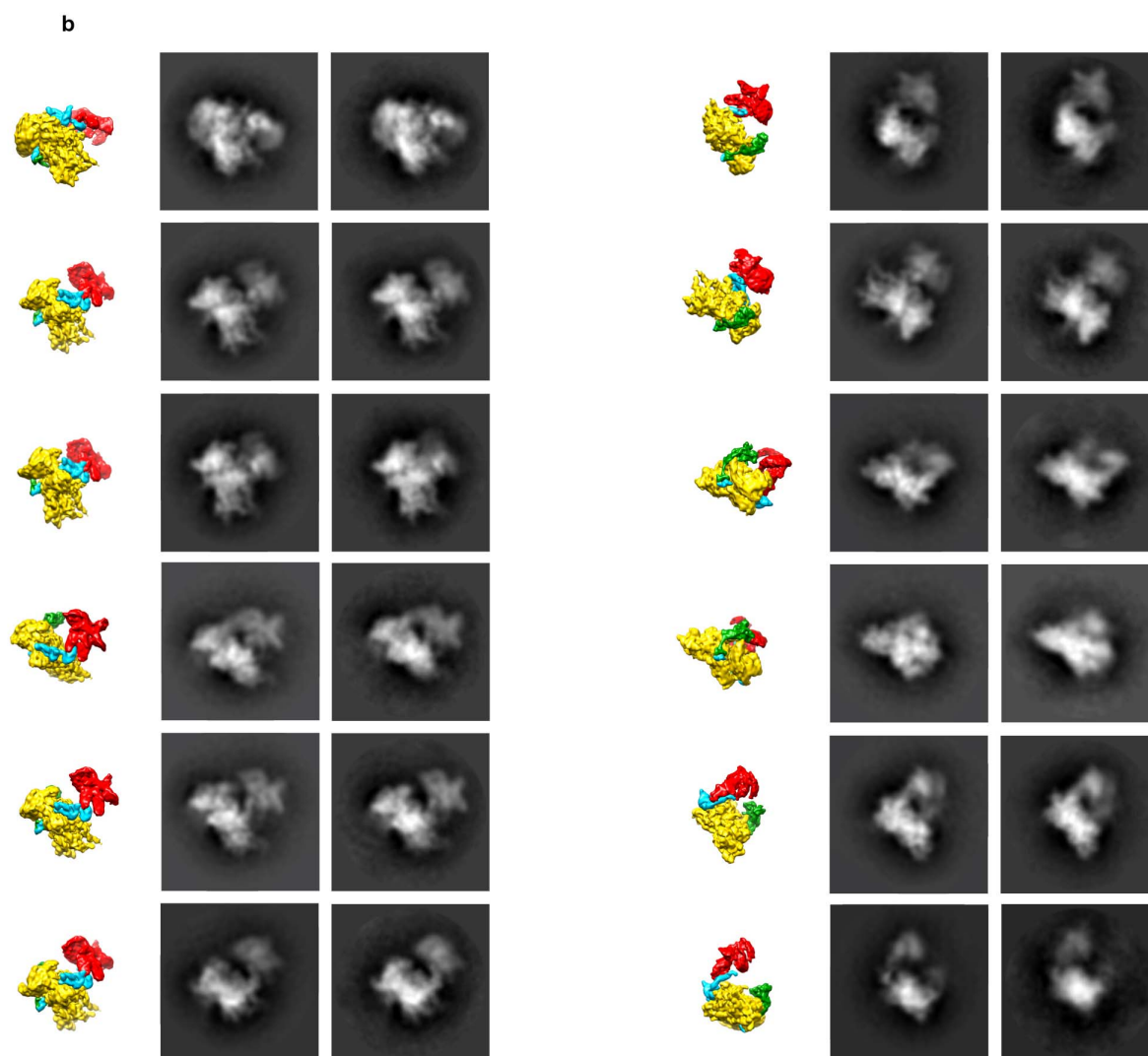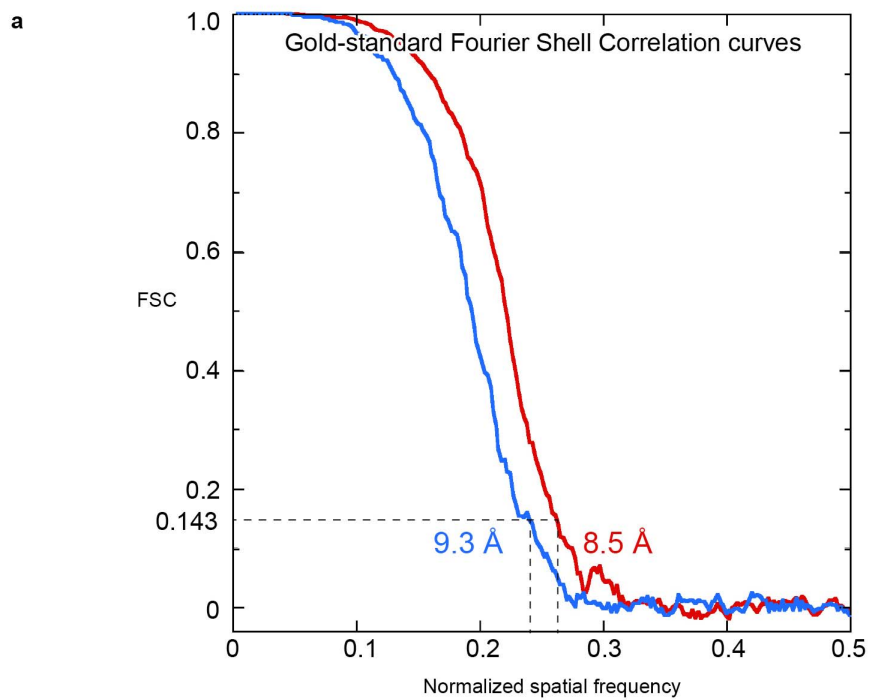
Extended Data Figure 4 | Measured resolution and reference-free two-dimensional classification of IRES-bound ribosomal complexes. a, Gold-standard Fourier shell correlation (FSC) curves of the cryo-electron microscopy reconstruction of classes 2 (red line) and 4 (blue line) (also see Extended Data Fig. 3) indicating their estimated resolution. **b**, Right column on each side, two-dimensional classes obtained by reference-free classification of particles corresponding to 40S–eIF3–DHX29–ΔII-IRES complexes (class 4 in Extended Data Fig. 3). Middle column on each side, projection views of the class 4 cryo-electron microscopy map corresponding to the two-dimensional classes. Right column on each side, corresponding views of the segmented three-dimensional map coloured as in Fig. 1.
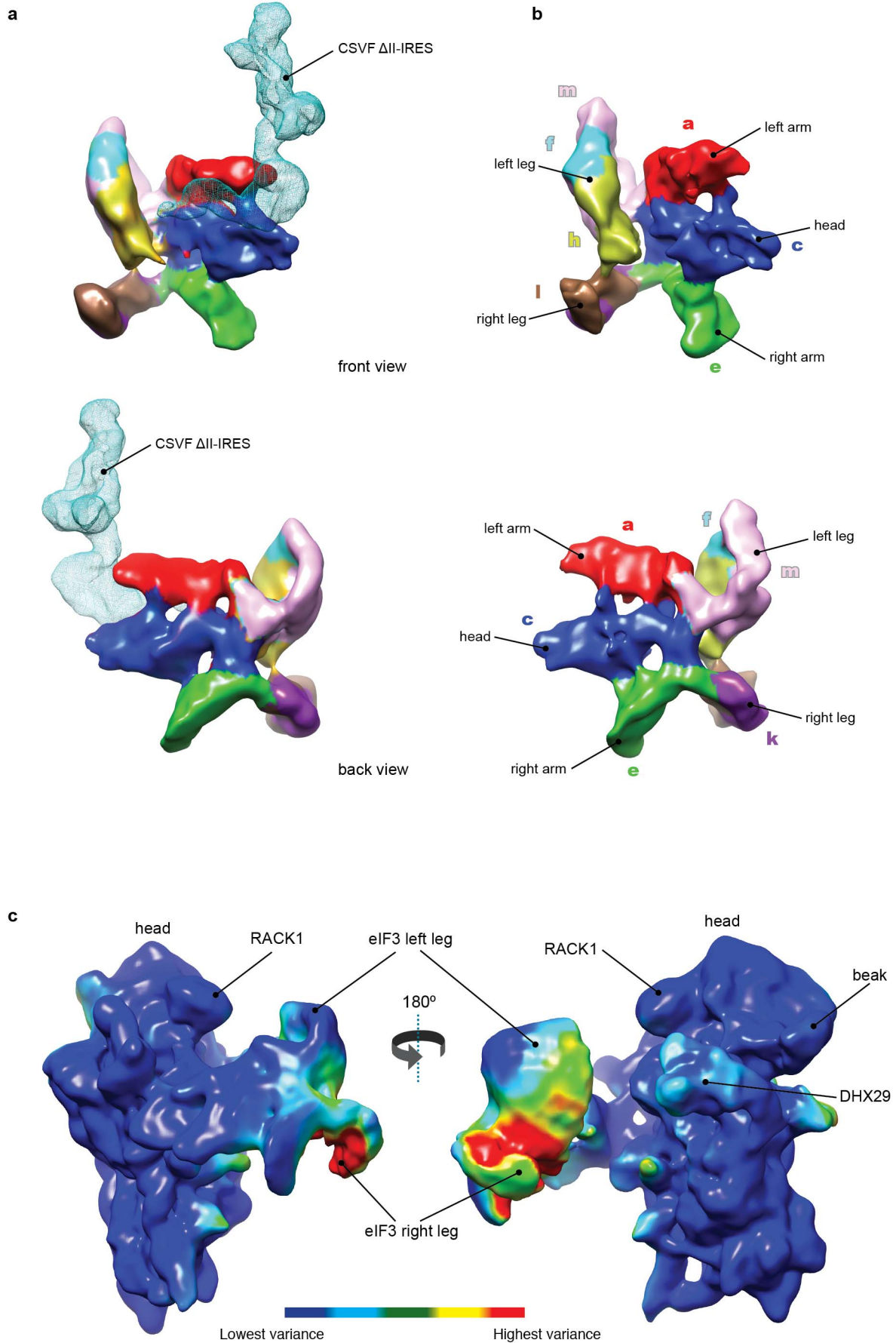
Extended Data Figure 5 | Correspondence between individual subunits and anthropomorphic features of the eIF3 core complex and three-dimensional variance of the 40S–DHX29–ΔII-IRES–eIF3 map. a, b, Front (upper panels) and back views (lower panels) of cryo-electron microscopy reconstructions of eIF3 as it appears in class 4 of the CSFV ΔII-IRES–40S–DHX29–eIF3 complex bound to the CSFV ΔII-IRES (**a**) and alone[13] (**b**), labelled to show anthropomorphic terms[12] and the localization of individual subunits in the core complex[11,24]. **c,** three-dimensional variance of the class 4 cryo-electron microscopy map, filtered to 20 Å, and coloured according to the computed three-dimensional variance (see Methods), from dark blue for the lowest variance to red for the highest variance. The map is filtered to the resolution at which the three-dimensional variance was estimated (~20 Å).
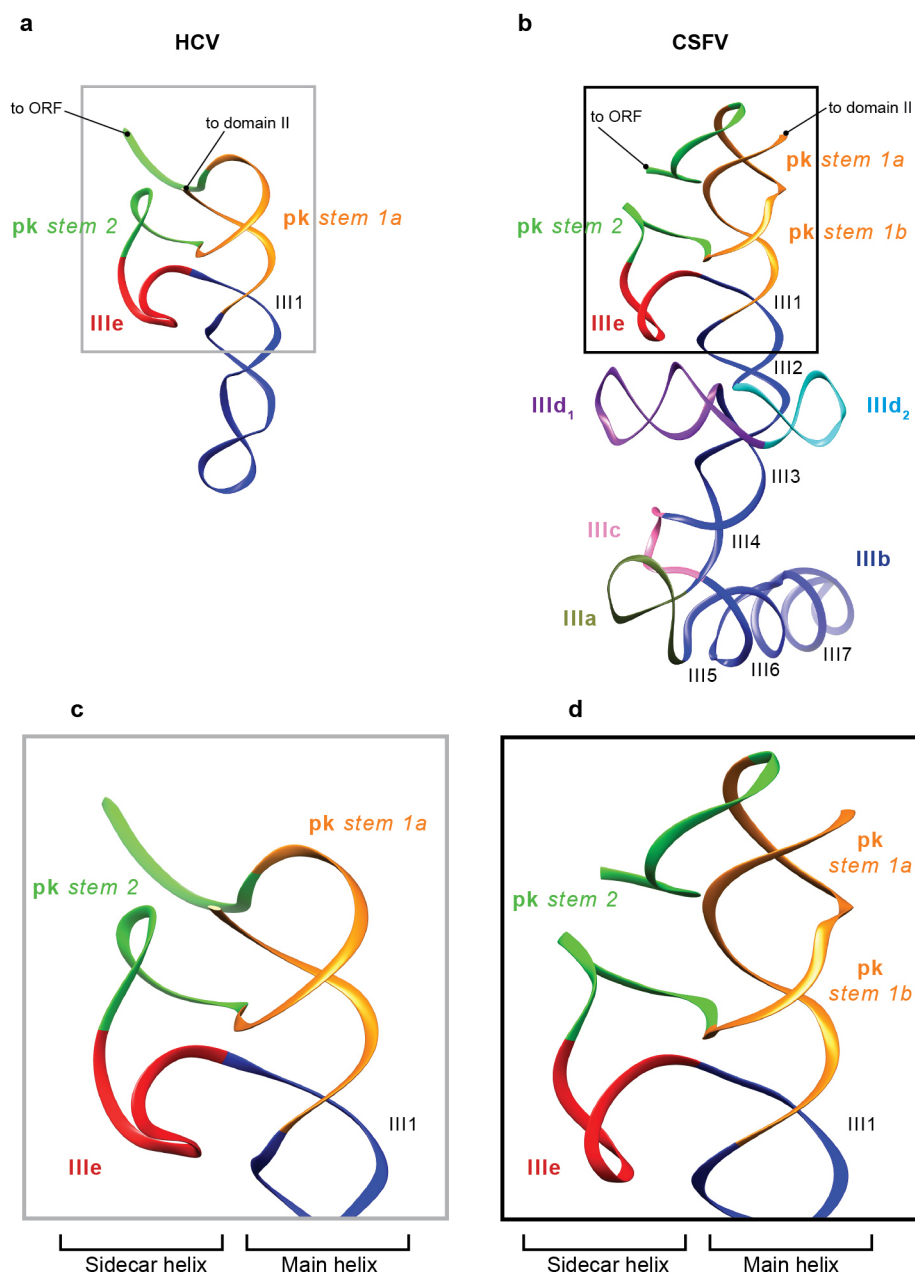
Extended Data Figure 6 | Comparison between the CSFV and the HCV pseudoknots. Views of the structures of the HCV pseudoknot, from the 3.6 Å resolution crystal structure, with an additional crystallization module extending from helix III₁[27] (**a**) (PDB: 3T4B) and the CSFV pseudoknot in the context of the 40S-subunit-bound ΔII-IRES (**b**) (this study) are shown in ribbon representation and coloured according to the scheme of the respective secondary structure diagrams (Extended Data Fig. 1a). **c, d,** Close-up views of HCV and CSFV pseudoknots, showing the 'main' helix, formed by helix III₁ and pseudoknot (pk) stem 1A (in HCV) or helix III₁, pk stem 1a and pk stem 1b (in CSFV), and the 'sidecar' helix, which contains subdomain IIIe, pk stem 2 and the two-base-pair helical segment of subdomain IIIf (see Extended Data Fig. 5a, b).
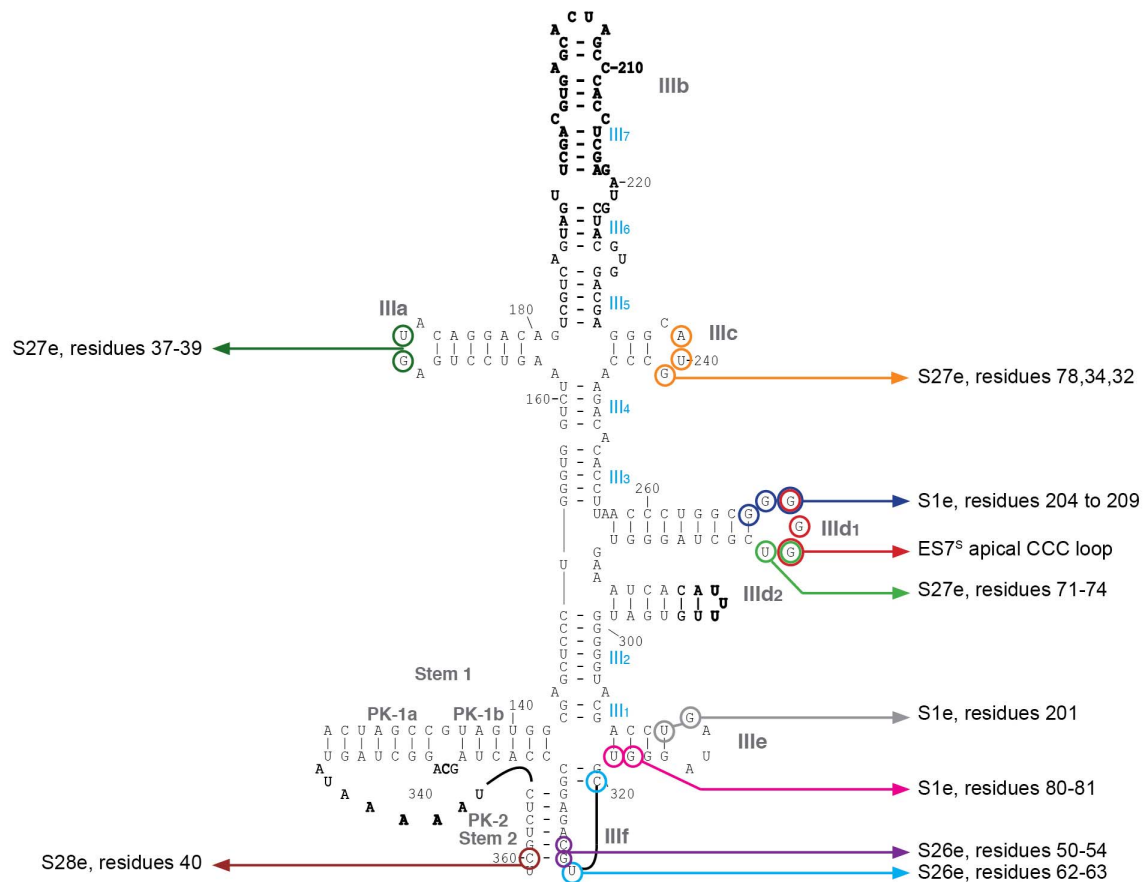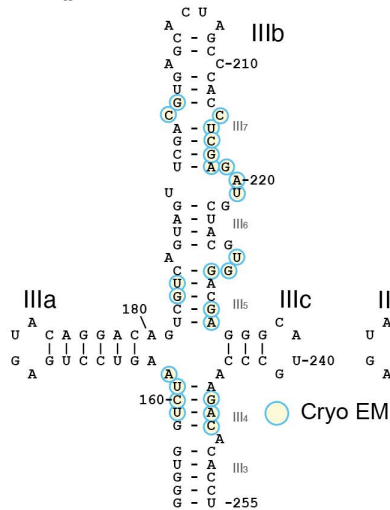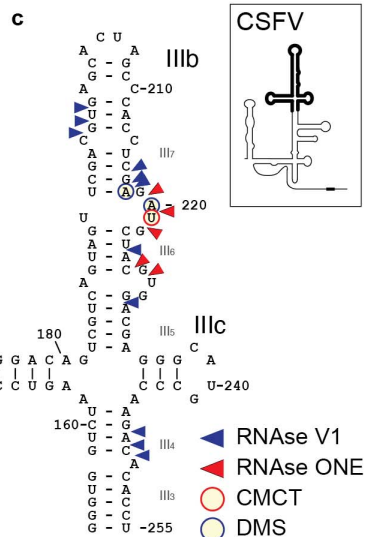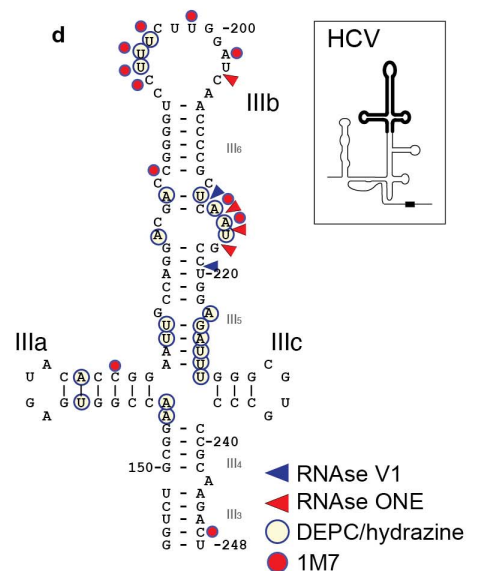
Extended Data Figure 7 | Molecular interactions of the CSFV ΔII-IRES with the 40S subunit and interactions of eIF3 with the HCV and CSFV IRESs. **a**, Secondary structure diagram of the CSFV ΔII-IRES, with nucleotides shown in different degrees of bold to show qualitatively their flexibility in the cryo-electron microscopy map (the more flexible, the bolder). Circled nucleotides interact with the indicated components of the 40S subunit. Ribosomal protein names and residue numbers are indicated according to the *Tetrahymena thermophila* 40S subunit[42]. **b–d**, Secondary structure diagram of the apical region of domain III of the CSFV IRES (**b**, **c**) and the HCV IRES (**d**). **b**, Contacts of eIF3 with the IRES in the cryoEM map of the 40S–ΔII-IRES–eIF3 complex.

c, **d**, Sites of strong protection of CSFV and HCV IRESs by native eIF3 from enzymatic cleavage and chemical modification, of protection of the HCV IRES by a 10-subunit form of eIF3 from 1M7 modification, or of interference with binding of eIF3 to the IRES by modification, as indicated in the keys[5,9,11]. Abbreviations: dimethyl sulphate (DMS), 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulphonate (CMCT), diethylpyrocarbonate (DEPC), 1-methyl-7-nitroisatoic anhydride (1M7). The inset panels show CSFV(**c**) and HCV IRESs (**d**), with helix III₄ and subdomains IIIa, IIIb and IIIc in bold.
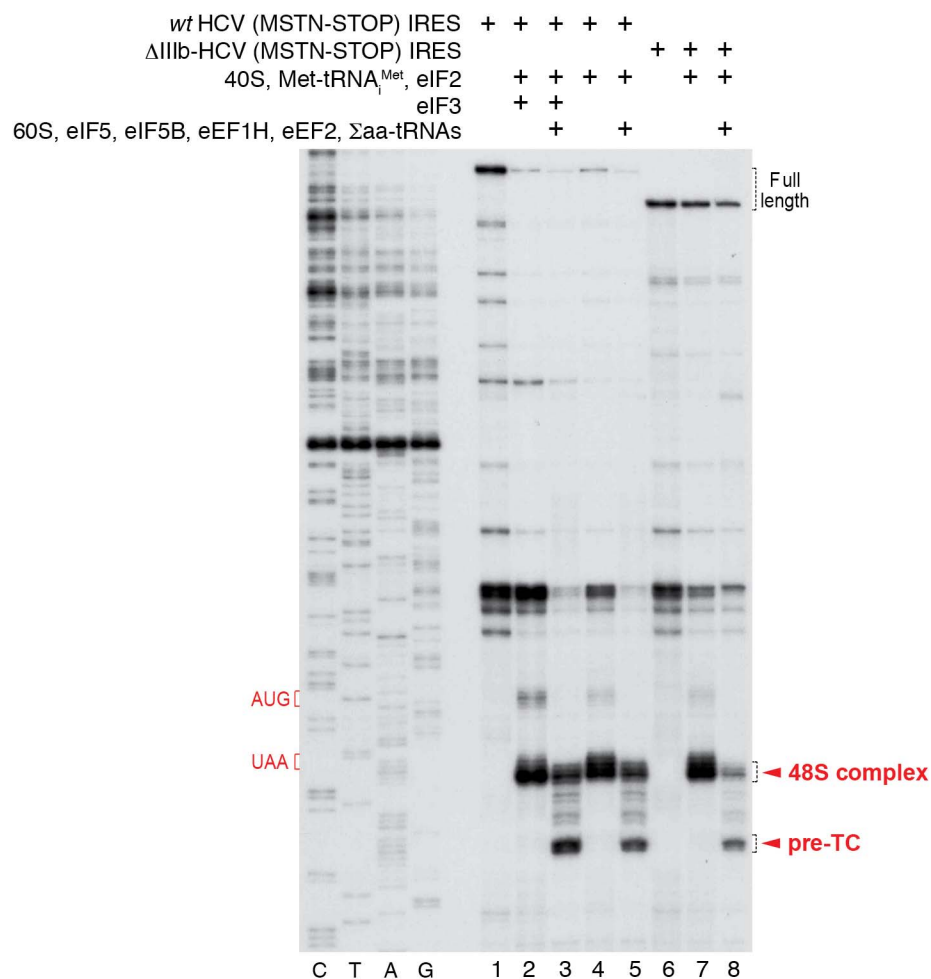
Extended Data Figure 8 | Formation of elongation-competent 80S ribosomes on the HCV IRES depending on the presence of eIF3. Toe-printing analysis of 48S initiation and 80S pre-termination complexes (pre-TC) assembled on the wild-type and ΔIIIb HCV (MSTN-STOP) mRNAs with translation components as indicated. The positions of the initiation and stop codons are shown on the left. Lanes C, T, A and G depict the cDNA sequence corresponding to the wild-type HCV (MSTN-STOP) mRNA. The gel reported in the figure is representative of results obtained from three technical replicates.
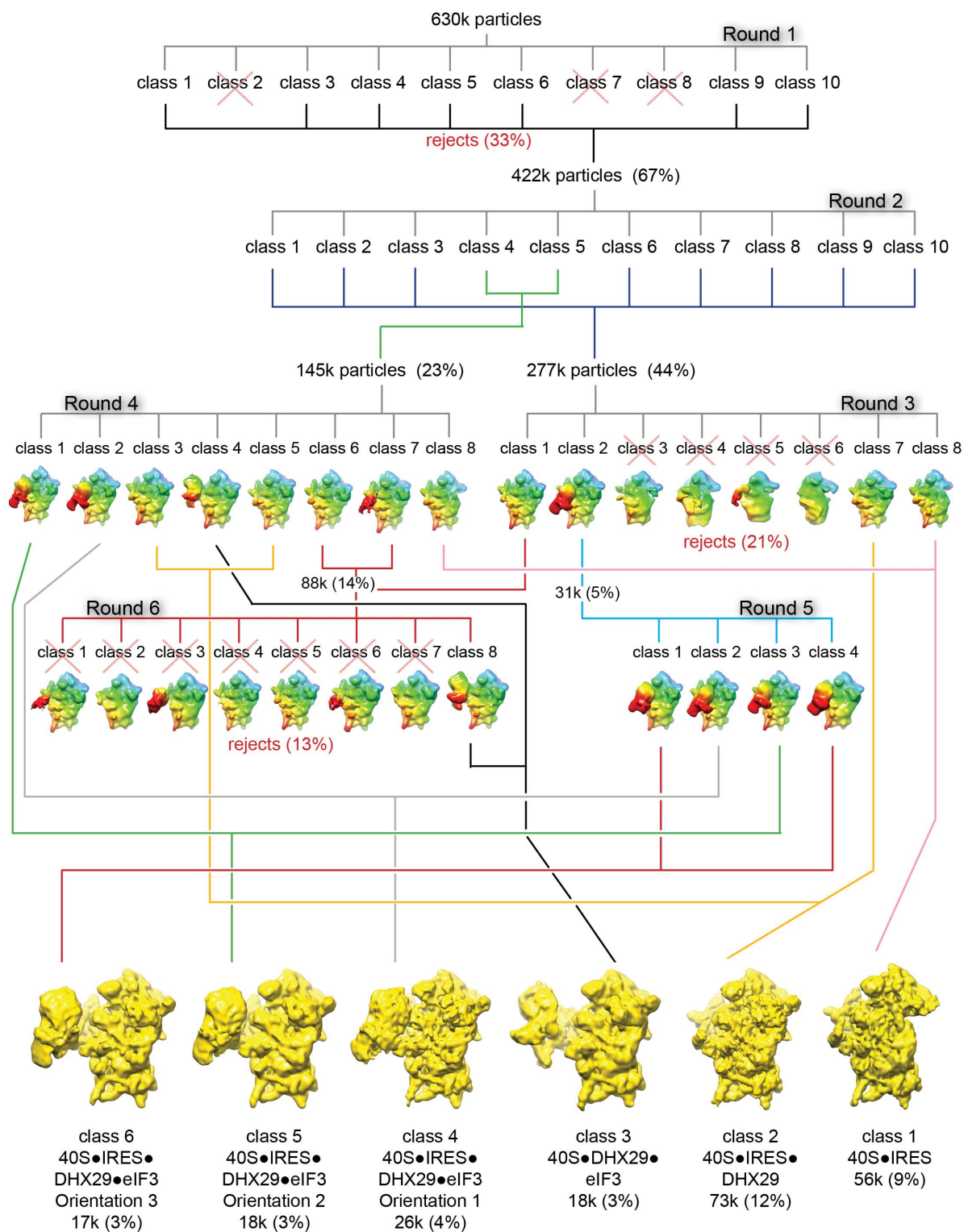
Extended Data Figure 9 | Unsupervised three-dimensional classification protocol. Details of the unsupervised three-dimensional classification. The classification included 6 rounds. For each round, the number of the particles included is indicated, as well as their percentages calculated over the full data set. The classes of rejected particles are crossed out in red and their percentages are indicated, also in red, as calculated over the full data set. Lines and brackets are drawn in different colours for clarity. Classes generated in rounds 3 to 6 are displayed and coloured by radial distance in Chimera UCSF[43] in order to help in the visual discrimination of differences in features among the classes.

Accelerated growth in the absence of DNA replication origins

Michelle Hawkins[1]*, Sunir Malla[2], Martin J. Blythe[2], Conrad A. Nieduszynski[1]* & Thorsten Allers[1]*

DNA replication initiates at defined sites called origins, which serve as binding sites for initiator proteins that recruit the replicative machinery. Origins differ in number and structure across the three domains of life[1] and their properties determine the dynamics of chromosome replication. Bacteria and some archaea replicate from single origins, whereas most archaea and all eukaryotes replicate using multiple origins. Initiation mechanisms that rely on homologous recombination operate in some viruses. Here we show that such mechanisms also operate in archaea. We use deep sequencing to study replication in *Haloferax volcanii* and identify four chromosomal origins of differing activity. Deletion of individual origins results in perturbed replication dynamics and reduced growth. However, a strain lacking all origins has no apparent defects and grows significantly faster than wild type. Origin-less cells initiate replication at dispersed sites rather than at discrete origins and have an absolute requirement for the recombinase RadA, unlike strains lacking individual origins. Our results demonstrate that homologous recombination alone can efficiently initiate the replication of an entire cellular genome. This raises the question of what purpose replication origins serve and why they have evolved.

H. volcanii is a genetically tractable archaeon[2,3]; its 2.85 megabase main chromosome is replicated from several origins[4] using machinery homologous to that found in eukaryotes[1]. To characterize replication dynamics in *H. volcanii*, we generated replication profiles by deep sequencing the wild isolate DS2 and laboratory strain H26 (Supplementary Table 1). Read counts from asynchronous replicating cells were normalized to non-replicating cells (Extended Data Fig. 1)[5]. Peaks in relative copy number correspond to sequences that are over-represented in replicating cells and therefore identify active origins (Fig. 1). In the wild isolate DS2, peaks at 0 and 1593 kb of the main chromosome co-localize with previously described origins (*oriC1* and *oriC2*, respectively), as do peaks in the mega-plasmid profiles (Extended Data Fig. 2)[4]. The peak at 571 kb represents a third chromosomal origin, *oriC3* (Extended Data Fig. 3). Unlike *oriC1* and *oriC2*, *oriC3* is not situated at a nucleotide skew inflection point[4]; in bacteria and archaea, such inflection points reflect origin use over evolutionary timescales[6]. This is consistent with infrequent use of *oriC3* or the recent acquisition of an origin at this location.

The sharp peaks reflect discrete origins, whereas the smooth valleys represent broad zones of termination[7]. Broad termination zones (as opposed to specific termination sites) have been described in other archaea[8] and in eukaryotes[7,9], suggesting they are a feature of chromosomes with multiple origins. The variable peak heights indicate that the chromosomal origins differ in activity; this interpretation is supported
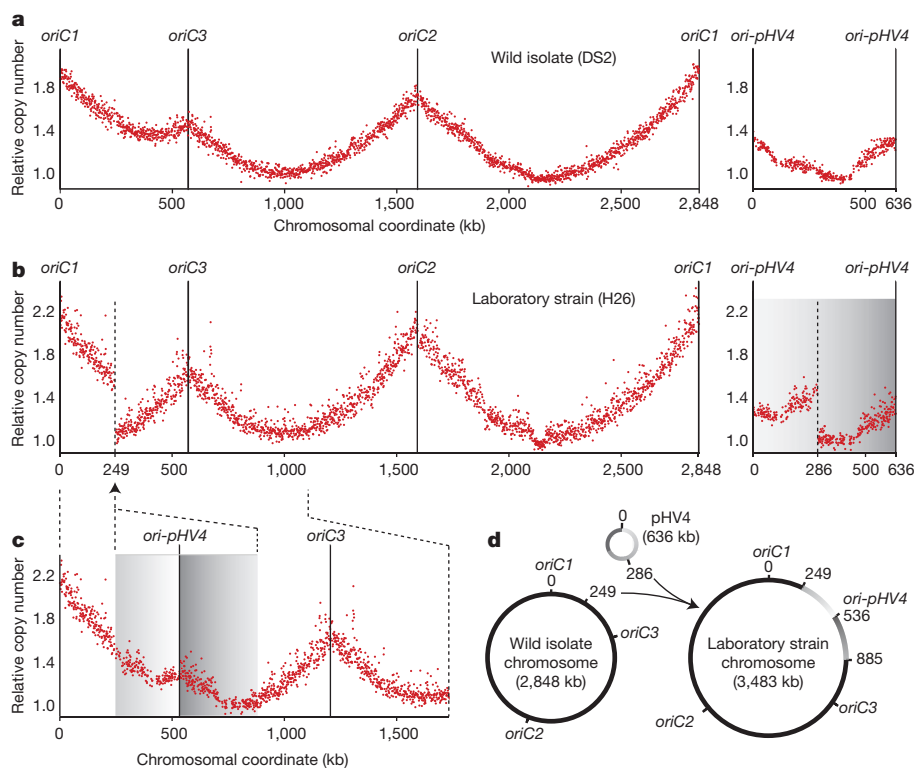
Figure 1 | Replication profiles for *H. volcanii* wild isolate and laboratory strain. a, Relative copy number plotted against chromosomal coordinate for the main chromosome and pHV4 of wild isolate DS2. Circular chromosomes are shown linearized at position 0, vertical lines mark replication origins. **b**, Sequence reads for laboratory strain H26 mapped to the reference genome (DS2). pHV4 shading reflects chromosomal coordinate. **c,** Sequence reads for H26 mapped to a reconstructed assembly of the main chromosome with pHV4 integrated at approximately 249 kb. Grey shading from **b** indicates the orientation of pHV4 integration. **d**, Integration of pHV4 into the main chromosome of H26.

[1]School of Biology, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK. [2]Deep Seq, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK.
*These authors contributed equally to this work.

by mathematical modelling (A. de Moura, personal communication) and plasmid-based assays (Extended Data Fig. 3c). Such a functional hierarchy of origins may be due to different use and/or activation times[9,10].

Laboratory strain H26 shows discontinuities in the replication profiles of the main chromosome and mega-plasmid pHV4 (at 249 and 286 kb respectively; Fig. 1b). Discontinuities indicate substantial differences in replication time between adjacent regions and suggest genome rearrangements[11]. We determined this rearrangement to be integration of pHV4 into the main chromosome (Fig. 1c, d and Extended Data Fig. 4). Remapping the data to a reconstructed genome sequence results in a continuous profile (Fig. 1c). The extra peak at 535 kb corresponds to the integrated pHV4 origin, ori-pHV4.

If an origin is active in all cells and used only once per generation, the ratio of origin to terminus regions cannot exceed 2:1; values greater than 2:1 are only possible if concurrent rounds of replication are initiated. The ratio of the wild isolate is 2:1 (Fig. 1a), but exceeds 2:1 for H26 (Fig. 1c). This is consistent with concurrent rounds of replication and precludes the existence of alternating phases of replication and segregation in H. volcanii (in contrast to eukaryotes and crenarchaea such as Sulfolobus[12]). Therefore, regulated timing of origin activation is unlikely and the peak height differences we observe are probably due to differences in origin use.

We tested the requirement for origins by chromosomal deletion in strain H26 (hereafter designated wild type). All combinations of origin deletion resulted in viable strains, including a strain deleted for all four chromosomal origins (Fig. 2a). Deletion of individual origins led to minor changes in DNA content (notably ΔoriC3), but the strain lacking all chromosomal origins had a DNA content profile indistinguishable from wild type (Fig. 2b). We used pairwise growth competition to quantify strain fitness (Fig. 2c). Single origin deletion strains grew

slower than wild type, with strains lacking oriC3 showing the greatest growth defect. Surprisingly, the strain deleted for all four origins grew 7.5% faster than wild type, and the strain lacking the three most active origins (oriC1,2,3) grew 5.5% faster (Fig. 2c). In fact, growth rate correlates inversely with the activity of remaining origins. For example, the ΔoriC2,3 strain retains the most active origin oriC1 and has a 0.8% growth defect, whereas the ΔoriC1,2 strain has lost the two most active origins and has a 2.3% growth advantage.

How could genome replication be maintained despite the deletion of all chromosomal origins? Deletion of known origins may reveal dormant origins as seen in yeast[13,14]. Alternatively, replication could initiate independently of canonical origins, with little or no site specificity. To distinguish between these possibilities, we profiled replication in the origin deletion strains (Fig. 3 and Supplementary Table 1). The peaks associated with deleted origin(s) are no longer evident and there are no new discrete peaks. The minima have relocated indicating that there are no enforced termination sites. The profiles of strains deleted for all (or the three most active) origins show a zone of copy number enrichment near the ΔoriC2 locus (Fig. 3e, f; in the region of 2230 kb). However, this does not resemble the sharp peaks associated with characterized origins (Figs 1 and 3a). Therefore, we find no evidence for activation of dormant origins.

Instead, the profiles are consistent with origin-independent initiation. In contrast to the sharp peaks observed in the wild type, profiles of the single origin deletion strains show global flattening that has rounded the remaining peaks (Fig. 3b–d); the minima at termination zones are also shallower. Sharp peaks indicate discrete origin sites, therefore peak flattening is a consequence of replication initiation at dispersed sites. The profiles of strains deleted for all (or the most active) origins are largely flat, consistent with widespread origin-independent initiation (Fig. 3e, f).

We considered two mechanisms for dispersed initiation. Origins are binding sites for the initiator protein ORC1; in the absence of origins,
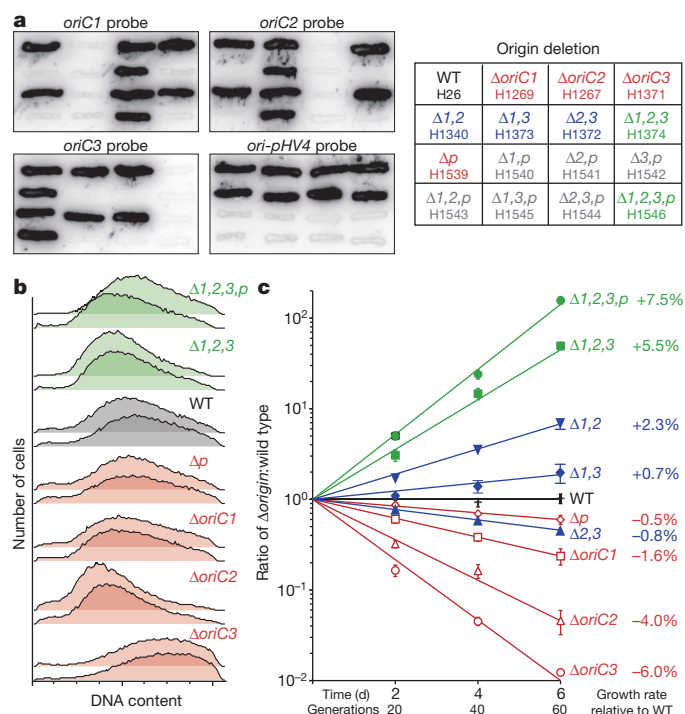
Figure 2 | Characterization of origin deletion strains. a, Deletion strains were confirmed by hybridization with origin-specific probes (WT, wild type; 'p', ori-pHV4). **b**, Flow cytometry was used to measure DNA content of origin deletion strains, biological replicates are shown; no differences in cell size were observed (data not shown). **c**, Pairwise growth competition assays comparing wild-type (H54, bgaHa+) and origin deletion strains. The average and standard error of four independent replicates are plotted.
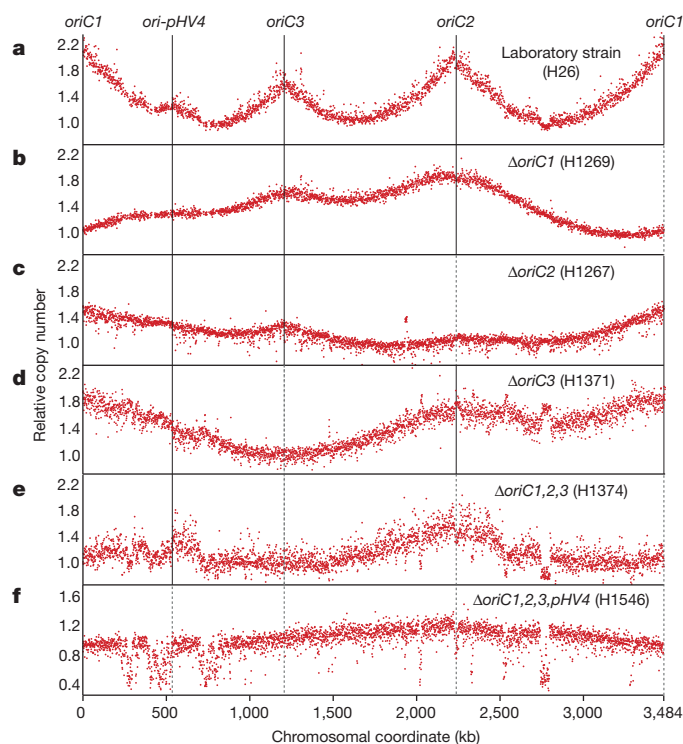
Figure 3 | Origin deletion strain replication profiles. Comparison of replication profiles for (**a**) laboratory strain (wild type), (**b**) ΔoriC1, (**c**) ΔoriC2, (**d**) ΔoriC3, (**e**) ΔoriC1,2,3 and (**f**) ΔoriC1,2,3,pHV4 mutants. Relative copy number for the main chromosome with integrated pHV4 was derived and shown as in Fig. 1c; dashed lines mark the location of deleted origins.
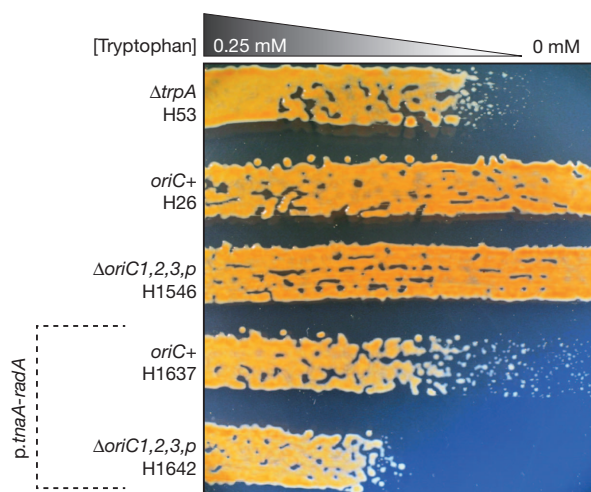
Figure 4 | RadA recombinase is essential in an *ΔoriC1,2,3,pHV4* mutant. *radA* was placed under control of the tryptophan-inducible p.*tnaA* promoter, in *oriC⁺* and *ΔoriC1,2,3,pHV4* strains (H1637 and H1642). The former grows slowly in the absence of tryptophan whereas the latter is inviable. Absence of tryptophan does not affect the growth of *oriC⁺* and *ΔoriC1,2,3,pHV4* control strains (H26 and H1546); the *ΔtrpA* control strain (H53) is auxotrophic for tryptophan.

ORC1 could bind non-specifically throughout the genome[15]. Alternatively, dispersed initiation could rely upon homologous recombination. Origin-independent replication can occur when recombination (D-loop) or transcription (R-loop) intermediates are used to prime replication[16,17]. We note that in strains deleted for all or the most active origins, the zone of copy number enrichment in the region of 2230 kb is near the *rrnB* ribosomal RNA (rRNA) operon (Fig. 3e, f; 2234–2239 kb). Highly transcribed DNA is associated with elevated recombination levels[18], therefore D-loops and R-loops in the *rrnB* region could facilitate replication initiation. This is analogous to recombination-dependent replication in viruses[17] and to DNA damage-inducible replication in *Escherichia coli*. The latter is known as 'stable DNA replication' and occurs in the absence of *oriC* or the initiator protein DnaA; instead it relies on recombination catalysed by RecA to initiate replication[16].

H. volcanii mutants lacking RadA (the archaeal RecA/Rad51 homologue) are viable but defective in recombination. Unlike RecA, RadA does not have a secondary role in activating an SOS response[19]. However, RadA is essential for the replication of pHV2-based plasmids, which do not use ORC-based initiation[20]. We attempted to delete *radA* from the origin deletion strains using established methods[21]. This was successful in the wild-type and single origin deletion strains, but only a single *ΔoriC1,2,3 ΔradA* isolate was recovered; this strain had undergone a chromosomal rearrangement involving *ori-pHV4* (Extended Data Fig. 5). We were unable to delete *radA* from the strain lacking all four origins, indicating that recombination is essential in the absence of replication origins. To confirm this, we placed *radA* under control of a tryptophan-inducible promoter (Extended Data Fig. 6)[22]. In the absence of tryptophan, when this promoter is tightly repressed, wild-type cells with inducible *radA* are viable whereas origin-less cells fail to grow (Fig. 4).

Work by Kogoma[16] showed that *E. coli oriC* mutants can use homologous recombination to initiate replication. However, these cells show profound growth defects[23]. In contrast, origin-less strains of *H. volcanii* grow faster than wild type. Furthermore, recombination-dependent replication in *E. coli* is only possible in strains harbouring suppressor mutations (for example *sdrA*, which stabilizes R-loops by inactivating RNaseHI[16]). We found no mutations in any of the four *H. volcanii* RNaseH genes. Only five single nucleotide polymorphisms (SNPs) were identified in the strain lacking all origins, and all of these SNPs are already present in the respective parent strains (Extended Data

Table 1). Therefore, we found no evidence for suppressors, akin to those reported by Kogoma and colleagues[23], which are required for growth in the absence of origins.

Our results indicate that it is possible to replicate an entire genome by recombination-dependent initiation alone, with no apparent cost to fitness. How might this be accomplished? In wild type, binding of ORC1 at origins leads to recruitment of the replicative helicase MCM, which may be rate-limiting for initiation[1]. In the single origin deletion strains, liberation of MCM from deleted origins could stimulate recombination-dependent initiation, resulting in flattening of the replication profiles (Fig. 3b–d). We postulate that the activity of origins correlates with their affinity for MCM. Therefore, deleting an active origin (*oriC1*) liberates more MCM than deleting a weak origin (*oriC3*). This liberated MCM is recruited to D-loops and used to initiate recombination-dependent replication. Our observation that growth rate correlates inversely with the activity of remaining origins (Fig. 2c) suggests that recombination-dependent replication is more efficient than origin-dependent replication, but the former has a lower affinity for MCM. Consistent with this, *Pyrococcus abyssi* MCM is recruited both to the origin and to a region containing rRNA and transfer RNA (tRNA) genes; the latter becomes the main binding site in stationary phase, suggesting liberation of MCM from the origin[24].

What then is the purpose of replication origins? It is assumed that regularly spaced origins ensure genome duplication in the shortest possible time[25]. This assumption is challenged by our data showing that origin-less cells grow faster than wild type. Alternatively, defined origins can be used to coordinate the direction of replication with the orientation of highly expressed genes. Collisions between replication and transcription machineries can stall DNA replication, and restarting stalled forks by recombination entails a risk of genome rearrangements. We did not observe any such rearrangements, except when the *ΔoriC1,2,3* strain was challenged with inactivation of recombination (Extended Data Fig. 5). Moreover, the rapid growth of origin-less mutants suggests that collisions between replication and transcription are less problematical than assumed.

Regulated initiation at origins allows for coordination of genome replication with segregation. This is critical in organisms with tightly regulated ploidy, such as *E. coli*, *Sulfolobus* and most eukaryotes[12]. However, *H. volcanii* is highly polyploid, tolerates variation in genome copy number[26] and there is no evidence for a regulated cell cycle (I. Duggin, personal communication). We suggest that the high ploidy of *H. volcanii* enables the accelerated growth of origin-less strains, in contrast to the growth defects observed in *E. coli*[23]. With a ploidy of 20, *H. volcanii* can rely on stochastic partitioning to ensure that daughter cells inherit a genome complement. However, it is vital that these 20 genome sequences are equalized to prevent the accumulation of recessive mutations, and this requires efficient recombination. In yeast, a screen for gene deletions that are lethal in polyploid cells found that almost all such mutations affect genomic stability, notably by impairing recombination[27]. Therefore, polyploidy creates a situation (in yeast) where homologous recombination becomes essential; it follows that naturally polyploid organisms such as *H. volcanii* are heavily reliant upon recombination. Indeed, *radA* mutants of *H. volcanii* suffer a more severe growth defect than *recA* mutants of *E. coli*[20].

In *H. volcanii*, origin-dependent initiation of replication seems to offer no demonstrable advantage; however, cells lacking individual origins are disadvantaged. We propose that origins are selfish genetic elements that ensure their own replication. Over time, origins become integrated with cellular processes such as the cell cycle, to coordinate genome duplication, segregation and cell division; ultimately this results in reduced ploidy. Propagation of selfish elements within a population requires a sexual process and lateral gene transfer by cell mating has been observed in *H. volcanii*[28]. It is notable that most archaeal origins are adjacent to the gene for their cognate initiator protein ORC1 (ref. 1). Such tight linkage, which is typical of selfish elements, ensures that origins acquired by lateral gene transfer can successfully subvert the

replicative machinery of their host. This is known as the replicon take-over hypothesis, where the host cell chromosome becomes dependent on extra-chromosomal elements for its propagation[29]. The replicon takeover hypothesis has until now focused on the DNA replication apparatus, but our findings suggest that origins can also behave as selfish genetic elements.

METHODS SUMMARY

Strains, plasmids, oligonucleotides and probes are given in Extended Data Tables 2–4. *H. volcanii* was grown and genetically manipulated as described previously[30]. Pairwise growth competition assays were performed as described previously[21], except that wild-type and mutant strains were mixed in a 1:1 ratio. Pulsed-field gels and flow cytometry were performed as described previously[21,26]. Genomic DNA for deep sequencing was isolated from 100 ml culture in stationary phase (A650 > 1) or 1 litre in exponential phase (A650 ~ 0.1)[30]. Library preparation and sequencing was performed according to SOLiD instructions and analysed using custom Perl scripts[5].

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Robinson, N. P. & Bell, S. D. Origins of DNA replication in the three domains of life. *FEBS J.* **272,** 3757–3766 (2005).
2. Hartman, A. L. *et al.* The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE* **5,** e9605 (2010).
3. Leigh, J. A., Albers, S. V., Atomi, H. & Allers, T. Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. *FEMS Microbiol. Rev.* **35,** 577–608 (2011).
4. Norais, C. *et al.* Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet.* **3,** e77 (2007).
5. Müller, C. A. & Nieduszynski, C. A. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res.* **22,** 1953–1962 (2012).
6. Sernova, N. V. & Gelfand, M. S. Identification of replication origins in prokaryotic genomes. *Brief. Bioinform.* **9,** 376–391 (2008).
7. de Moura, A. P., Retkute, R., Hawkins, M. & Nieduszynski, C. A. Mathematical modelling of whole chromosome replication. *Nucleic Acids Res.* **38,** 5623–5633 (2010).
8. Duggin, I. G., Dubarry, N. & Bell, S. D. Replication termination and chromosome dimer resolution in the archaeon *Sulfolobus solfataricus*. *EMBO J.* **30,** 145–153 (2011).
9. Retkute, R., Nieduszynski, C. A. & de Moura, A. Dynamics of DNA replication in yeast. *Phys. Rev. Lett.* **107,** 068103 (2011).
10. Duggin, I. G., McCallum, S. A. & Bell, S. D. Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc. Natl Acad. Sci. USA* **105,** 16737–16742 (2008).
11. Skovgaard, O., Bak, M., Lobner-Olesen, A. & Tommerup, N. Genome-wide detection of chromosomal rearrangements, indels, and mutations in circular chromosomes by short read sequencing. *Genome Res.* **21,** 1388–1393 (2011).
12. Lindås, A. C. & Bernander, R. The cell cycle of archaea. *Nature Rev. Microbiol.* **11,** 627–638 (2013).
13. Dershowitz, A. *et al.* Linear derivatives of *Saccharomyces cerevisiae* chromosome III can be maintained in the absence of autonomously replicating sequence elements. *Mol. Cell. Biol.* **27,** 4652–4663 (2007).
14. Vujcic, M., Miller, C. A. & Kowalski, D. Activation of silent replication origins at autonomously replicating sequence elements near the HML locus in budding yeast. *Mol. Cell. Biol.* **19,** 6098–6109 (1999).
15. Vashee, S. *et al.* Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes Dev.* **17,** 1894–1908 (2003).
16. Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* **61,** 212–238 (1997).
17. Kreuzer, K. N. & Brister, J. R. Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. *Virol. J.* **7,** 358 (2010).
18. McGlynn, P., Savery, N. J. & Dillingham, M. S. The conflict between DNA replication and transcription. *Mol. Microbiol.* **85,** 12–20 (2012).
19. McCready, S. *et al.* UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium* sp. NRC-1. *Saline Syst.* **1,** 3 (2005).
20. Woods, W. G. & Dyall-Smith, M. L. Construction and analysis of a recombination-deficient (*radA*) mutant of *Haloferax volcanii*. *Mol. Microbiol.* **23,** 791–797 (1997).
21. Delmas, S., Shunburne, L., Ngo, H. P. & Allers, T. Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination. *PLoS Genet.* **5,** e1000552 (2009).
22. Large, A. *et al.* Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. *Mol. Microbiol.* **66,** 1092–1106 (2007).
23. Ogawa, T., Pickett, G. G., Kogoma, T. & Kornberg, A. RNase H confers specificity in the *dnaA*-dependent initiation of replication at the unique origin of the *Escherichia coli* chromosome *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **81,** 1040–1044 (1984).
24. Matsunaga, F. *et al.* Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in *Pyrococcus* sp. *Nucleic Acids Res.* **35,** 3214–3222 (2007).
25. Newman, T. J., Mamun, M. A., Nieduszynski, C. A. & Blow, J. J. Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkt728 (19 August 2013).
26. Breuert, S., Allers, T., Spohn, G. & Soppa, J. Regulated polyploidy in halophilic archaea. *PLoS ONE* **1,** e92 (2006).
27. Storchova, Z. *et al.* Genome-wide genetic analysis of polyploidy in yeast. *Nature* **443,** 541–547 (2006).
28. Rosenshine, I., Tchelet, R. & Mevarech, M. The mechanism of DNA transfer in the mating system of an archaebacterium. *Science* **245,** 1387–1389 (1989).
29. McGeoch, A. T. & Bell, S. D. Extra-chromosomal elements and the evolution of cellular DNA replication machineries. *Nature Rev. Mol. Cell Biol.* **9,** 569–574 (2008).
30. Allers, T., Ngo, H. P., Mevarech, M. & Lloyd, R. G. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* **70,** 943–953 (2004).

METHODS

Reagents. Strains, plasmids, oligonucleotides and probes are given in Extended Data Tables 2–4. *H. volcanii* was grown as described previously[30]. Pairwise growth competition assays were performed as described previously[21], except that wild-type and mutant strains were mixed in a 1:1 ratio (for further details see source data for Fig. 2c). Tryptophan gradient agar plates[31] were cast from a tapered wedge of Hv-Ca agar[30] containing 0.25 mM tryptophan, which was overlaid with a converse wedge of Hv-Ca agar.

Molecular genetic methods. Transformation of *H. volcanii* and genomic deletions were performed as described previously[30]. Standard molecular techniques were used; pulsed-field gels were performed as described previously[21]. Genomic DNA for deep sequencing was isolated from 100 ml Hv-YPC culture in stationary phase (A650 > 1) or 1 litre in exponential phase (A650 ~ 0.1) as described previously[30], followed by phenol:chloroform extraction. For flow cytometry, live cells in exponential phase (A650 ~ 0.1) were stained with acridine orange and immediately analysed using an Apogee A40 as described previously[26,32]; 50,000 cells were counted, doublet signals were removed by gating on peak/area plots and data analysed using FlowJo (TreeStar).

SOLiD sequencing and data analysis. Library preparation and sequencing was performed by Deep Seq (University of Nottingham) according to SOLiD instructions. Sequence reads were mapped to the *H. volcanii* genome (accession numbers CP001953–CP001957) using BioScope (version 1.3.1). Custom Perl scripts were used to calculate and plot replication profiles as described previously[5]. Deep sequencing data are available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41961.

Identification of SNPs. SNPs present at greater than 50% prevalence and with a coverage of more than 10 times (mean genome-wide coverage is approximately 150 times) are shown in Extended Data Table 1. Five SNPs were identified in the strain lacking all chromosomal origins (H1546), four of which are already present in the $oriC^+$ parent strain (H53). The remaining SNP leads to a predicted glycine to valine change in the hypothetical protein HVO_A0627; this mutation is already present in the $\Delta oriC1,2$ parent (H1340). However, it is absent from its $\Delta oriC1::trpA+ \Delta oriC2$ parent (H1293) and both these strains grow at near-identical rates (source data for Fig. 2c).

Isolation of *oriC3* by genetic screen for autonomously replicating sequences. We previously showed that genetic screens in *H. volcanii* isolate a single origin at a time, and that this can be circumvented by using origin deletion mutants[4]. Therefore, we deleted *oriC1* in a *Δori-pHV1* background, so that genetic screens would not be dominated by these two origins. Note that *ori-pHV1* was previously named *ori-pHV1/4*; this sequence hybridizes to two bands of about 690 kb and 86 kb on a pulsed-field gel, which correspond in size to pHV4 and pHV1, respectively[4]. However, it is now clear that pHV4 has integrated on the main chromosome in laboratory strains, therefore *ori-pHV1/4* cannot be present on pHV4. Consequently, we have renamed this origin as *ori-pHV1* and renamed *ori-pHV4-2* as *ori-pHV4*. To delete *oriC1*, the EcoRI–BspEI *oriC1* duplex unwinding element (DUE) fragment of pTA441 was replaced with the BamHI–XbaI $hdrB^+$ selectable marker from pTA187 to generate the *oriC1* deletion construct pTA946. Strain H300 was transformed with pTA946 as described previously[30], to generate the $\Delta oriC1::hdrB^+$ mutant H1023. Genomic DNA from strain H1023 was prepared as described previously[30]; 25 μg were partly digested with 0.5 units per microgram of AciI for 30 min and fragments of 4–8 kb were ligated in the ClaI site of plasmid pTA131. One microgram of this genomic library was used to transform the recombination-deficient strain H112; plasmid DNAs from six transformants were passaged through *E. coli* and sequenced. All six clones contained the autonomously replicating *oriC3* region (pTA1100; Extended Data Fig. 3).

Identifying the integration of pHV4 into the main chromosome. Genomic DNA from wild isolate DS2 and laboratory strain H26 was digested with ClaI, KpnI or NarI (Extended Data Fig. 4). A Southern blot was probed with PCR products upstream (US; primers RFB5F and RFB3R) and downstream (DS; primers RFBF and RFBR) of the H26 profile discontinuity (Fig. 1b). The upstream 3,646 base pair (bp) NarI fragment and downstream 7,478 bp KpnI fragment were isolated from H26 genomic libraries and cloned in pBluescript II SK+. The upstream clone pTA1238 and downstream clone pTA1236 contained chromosomal and pHV4 sequences (shown in Extended Data Fig. 4b), indicating that the entire 690 kb pHV4 had integrated into the main chromosome, by recombination between ISH18 insertion sequence elements HVO_0278 (chromosome) and HVO_A0279 (pHV4), as shown in Fig. 1d.

Deletion of *radA*. Deletion of *radA* was performed as described previously[21]. Briefly, the $\Delta radA::trpA^+$ construct pTA324 was used for chromosomal deletion of *radA* as described previously[30], but in the presence of pTA411 for *in trans* complementation of *radA* to facilitate efficient homologous recombination. Deletion of *radA* results in slow growth, the fraction of slow-growing colonies (*ΔradA* candidates) that proved to be *ΔradA* was 94%, 100%, 43% and 12% for the wild-type, *ΔoriC1*, *ΔoriC2* and *ΔoriC3* strains, respectively. Only a single *ΔradA ΔoriC1 ΔoriC2 ΔoriC3* colony (1 of 70 screened) was recovered; this strain had undergone a chromosomal rearrangement involving the part of integrated pHV4 containing *ori-pHV4* (Extended Data Fig. 5). We were unable to delete *radA* from the strain lacking all chromosomal origins (0 of 455 slow-growing colonies screened).

Generating tryptophan-inducible *radA* strains. Plasmid pTA1343 carries a recombinant *radA* allele under control of the tryptophan-inducible p.*tnaA* promoter[22]. The *radA* gene was cloned downstream of the p.*tnaA* promoter in pTA927 (ref. 33), from which a cassette comprising the t.L11e terminator, p.*tnaA* promoter, *radA* and t.Syn terminator was excised and linked to the *hdrB* marker from pTA187 (ref. 30), whereupon it was inserted between the upstream and downstream flanking regions of *radA* in pTA131 (ref. 30) to generate pTA1343 (Extended Data Fig. 6a). Further details are available upon request. pTA1343 was used to replace the native *radA* gene in H98 (wild type, generating H1637) and H1608 (*ΔoriC1,2,3,pHV4*, generating H1642) as described previously[30], except that transformants were plated on Hv-Ca + 5-FOA containing 0.25 mM tryptophan to ensure expression of the p.*tnaA*-*radA*$^+$ gene.

31. Bryson, V. & Szybalski, W. Microbial selection. *Science* **116,** 45–51 (1952).

32. Delmas, S., Duggin, I. G. & Allers, T. DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon *Haloferax volcanii. Mol. Microbiol.* **87,** 168–179 (2013).

33. Allers, T., Barak, S., Liddell, S., Wardell, K. & Mevarech, M. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii. Appl. Environ. Microbiol.* **76,** 1759–1769 (2010).

34. Alkan, C. *et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genet.* **41,** 1061–1067 (2009).

35. Pelve, E. A., Martens-Habbena, W., Stahl, D. A. & Bernander, R. Mapping of active replication origins *in vivo* in thaum- and euryarchaeal replicons. *Mol. Microbiol.* http://dx.doi.org/10.1111/mmi.12382 (16 September 2013).

36. Mullakhanbhai, M. F. & Larsen, H. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* **104,** 207–214 (1975).

37. Wendoloski, D., Ferrer, C. & Dyall-Smith, M. L. A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology* **147,** 959–964 (2001).

38. Lestini, R., Duan, Z. & Allers, T. The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in *Haloferax volcanii. DNA Repair* **9,** 994–1002 (2010).

a ABI SOLiD: bias from GC-content

Y = M0 + M1*x + ... M8*x^8 + M9*x^9	
M0	-53041
M1	3.8283e+5
M2	-7.0266e+5
M3	3.8303e+5
R	0.98295

b Read counts from stationary phase cells

c Corrected sequence reads vs. GC-content

d Read counts from stationary phase cells normalised for GC content
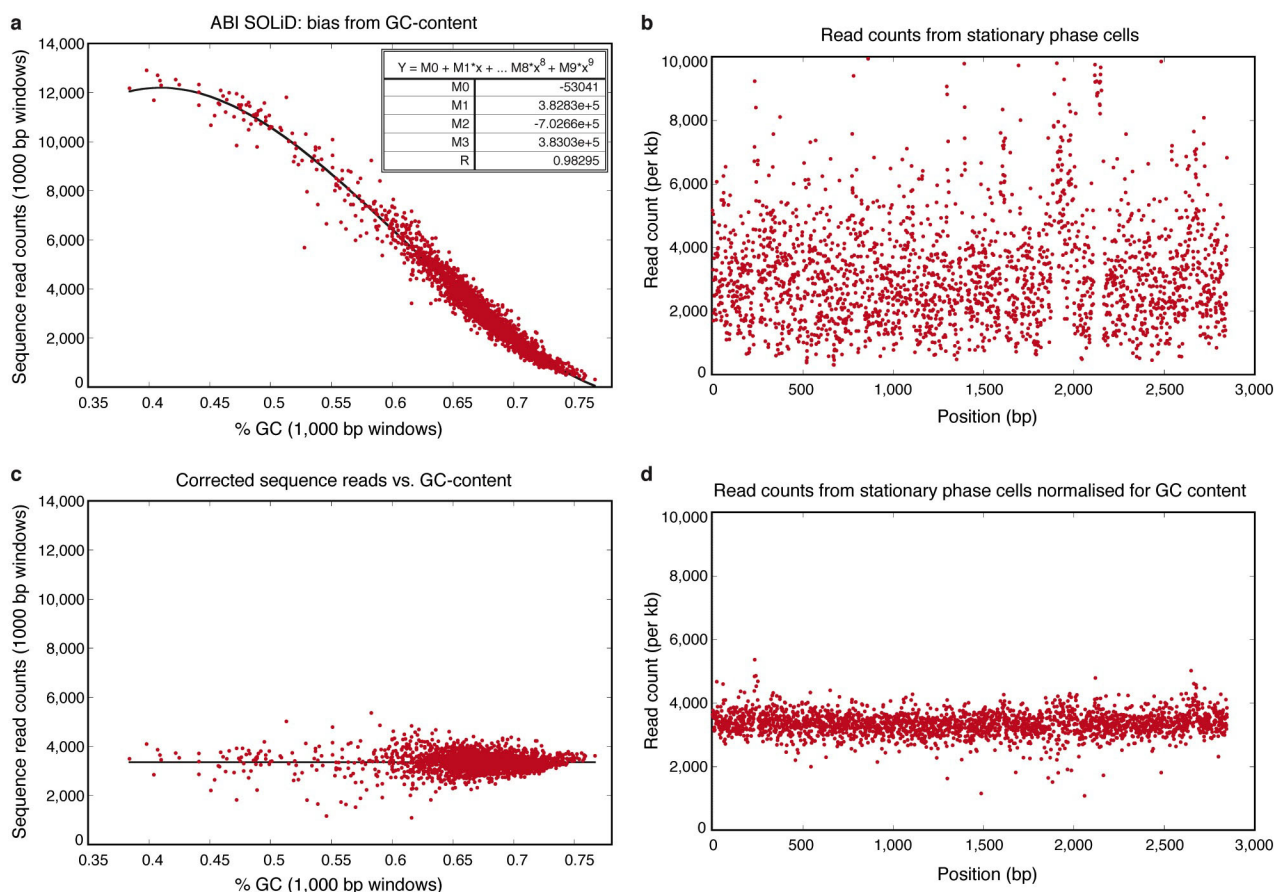
Extended Data Figure 1 | Correcting for GC-bias in deep sequencing data.
Sequence composition has previously been reported to influence the depth of
sequence coverage[34]. Therefore we investigated whether GC-content
contributes to the noise in our data. Sequence reads from the wild isolate (DS2)
stationary-phase sample were analysed with respect to GC-content. **a**, For each
1 kb window of unique sequence the number of mapped reads was plotted
against the GC-content of the window. We found a significant reduction in
mapped sequence reads at elevated GC-content. A polynomial equation (inset
and solid line) was fitted to the data. **b**, For each 1 kb window of unique

sequence, the read counts were plotted against chromosome position. **c**, Using
the method of Alkan *et al.*[34], we corrected for GC-bias using the polynomial
equation shown in **a** and then plotted the corrected sequence reads against
GC-content. **d**, GC-bias-corrected sequence reads are shown plotted against
chromosomal position. With no substantial continuing replication in the
stationary-phase sample, we can justify using this data set to normalize the
exponential phase data. Both normalization methods result in low noise
compared with studies that do not use a normalization step[35].
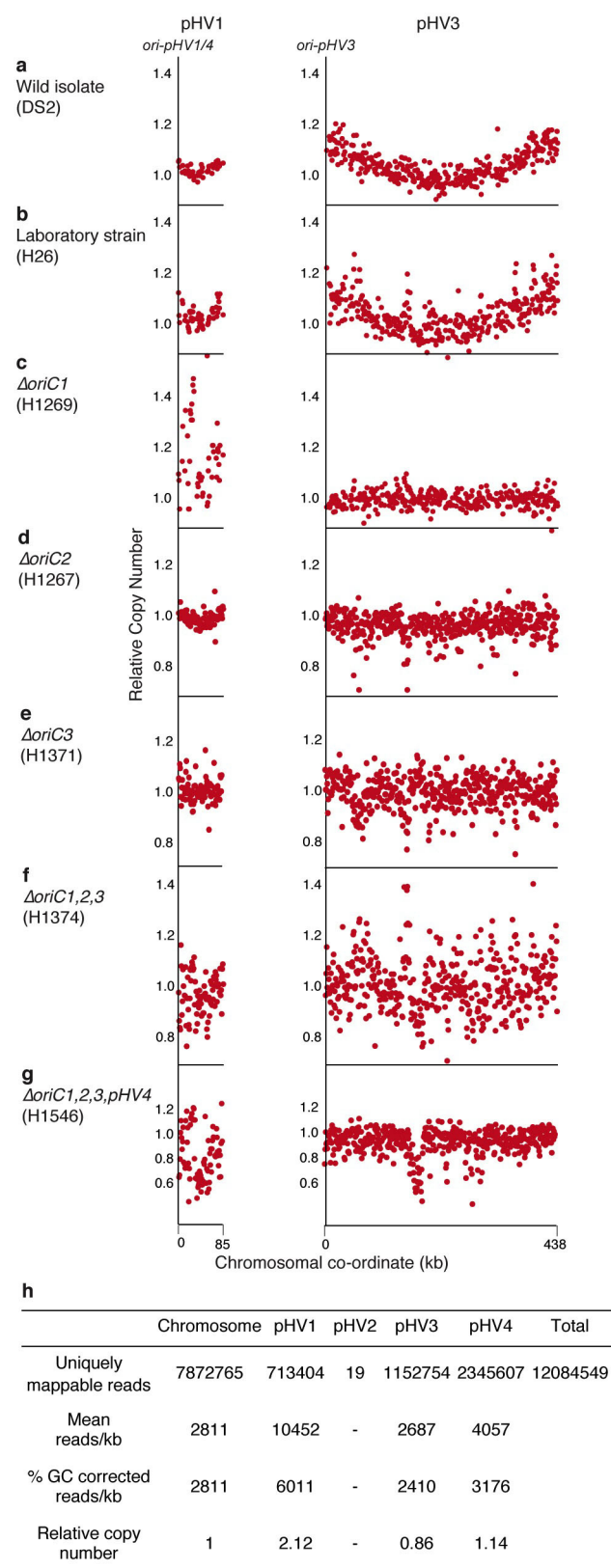
Extended Data Figure 2 | Replication profiles and copy numbers of mega-plasmids. Relative copy number plotted against chromosomal coordinate (kb) for pHV1 and pHV3 of (**a**) wild isolate DS2, (**b**) laboratory strain H26, (**c**) *ΔoriC1* H1269, (**d**) *ΔoriC2* H1267, (**e**) *ΔoriC3* H1371, (**f**) *ΔoriC1,2,3* H1374 and (**g**) *ΔoriC1,2,3,pHV4* H1546. Each mega-plasmid is shown linearized at position 0, the location of previously described origins[4]. The 6 kb pHV2 plasmid is not shown in the wild isolate owing to the scarcity of data points;

pHV2 is not present in laboratory strains (**b–g**). The pHV4 data for DS2 are shown in Fig. 1. Separate pHV4 data for laboratory strains (**b–g**) are excluded, because pHV4 is incorporated into the main chromosome in these strains (Figs 1 and 3). **h**, Relative copy number for each mega-plasmid was calculated using the GC-content normalized sequence counts from the stationary-phase data for laboratory strain H26.
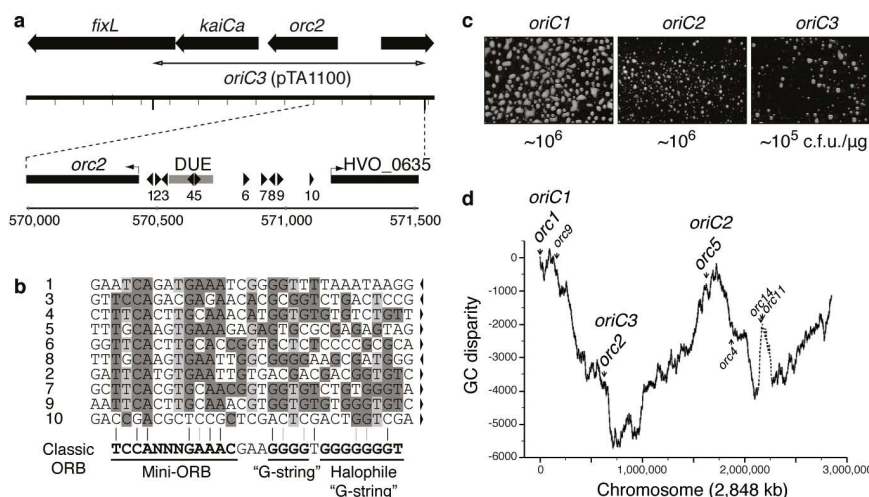
Extended Data Figure 3 | Characterization of *oriC3*. a, Sequence features of *oriC3*. Double-headed arrow indicates the autonomously replicating fragment recovered from a genomic library of H1023 (pTA1100; see Methods for details); solid arrows, open reading frames; triangles, repeats. The intergenic region upstream of *orc2* is typical of archaeal origins. It is enlarged to show the sequence features of *oriC3* duplex unwinding element (DUE). HVO_0635 encodes a conserved hypothetical protein. **b**, Sequence of intergenic repeats upstream of *orc2* (numbered in **a**, triangles show repeat orientation). Dark grey shading indicates match to consensus origin recognition box (ORB); bases conserved between repeats are indicated by light grey shading. **c**, Plasmid-based assays for the three chromosomal origins. Recombination-deficient strain H112 was transformed with 1 μg of pTA441 (*oriC1*), pTA612 (*oriC2*) or pTA1100 (*oriC3*). Transformants were plated with 100-fold dilution on Hv-Ca and incubated at 45 °C for 14 days. Numbers indicate transformation efficiency in colony-forming units (c.f.u.) per microgram of DNA. **d**, GC-disparity of main chromosome in wild isolate DS2 (adapted from ref. 4); positions of *orc* genes and replication origins are shown. The lack of a nucleotide disparity inflection point at *oriC3* suggests that this origin has been acquired recently or is used infrequently, consistent with the replication profile (Fig. 1a) and plasmid-based assay (Extended Data Fig. 3c).
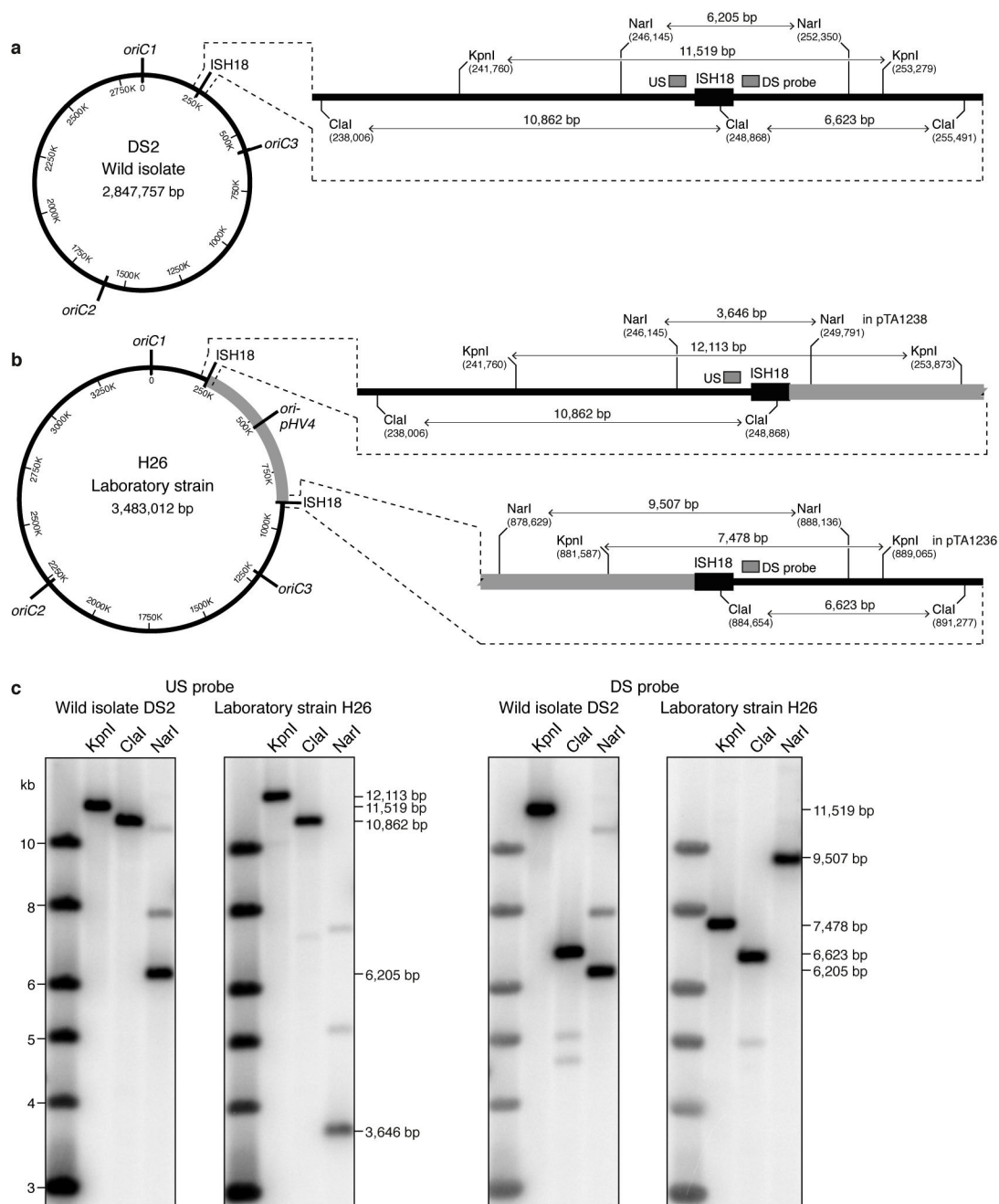
Extended Data Figure 4 | Identifying integration of pHV4 into the main chromosome. a, Map of region around ISH18 insertion sequence element HVO_0278 on the main chromosome of wild isolate DS2, showing restriction sites and probes used to determine the integration of pHV4. **b**, Map illustrating integration of pHV4 into the main chromosome of laboratory strain H26, by recombination between ISH18 HVO_0278 (chromosome) and ISH18 HVO_A0279 (pHV4). Regions upstream and downstream of the integration are depicted with the same restriction sites and probes shown in Extended Data Fig. 4a, in addition to the genomic fragments cloned in pTA1238 and pTA1236. **c**, Restriction fragment length polymorphisms in the main chromosome of laboratory strain H26. Genomic DNA from wild isolate DS2 and laboratory strain H26 was digested with KpnI, ClaI or NarI, and probed with sequences upstream (US) and downstream (DS) of ISH18 insertion sequence element HVO_0278. The upstream 3,646 bp NarI fragment of H26 was cloned in pTA1238, and the downstream 7,478 bp KpnI fragment of H26 was cloned in pTA1236. See Methods for details.
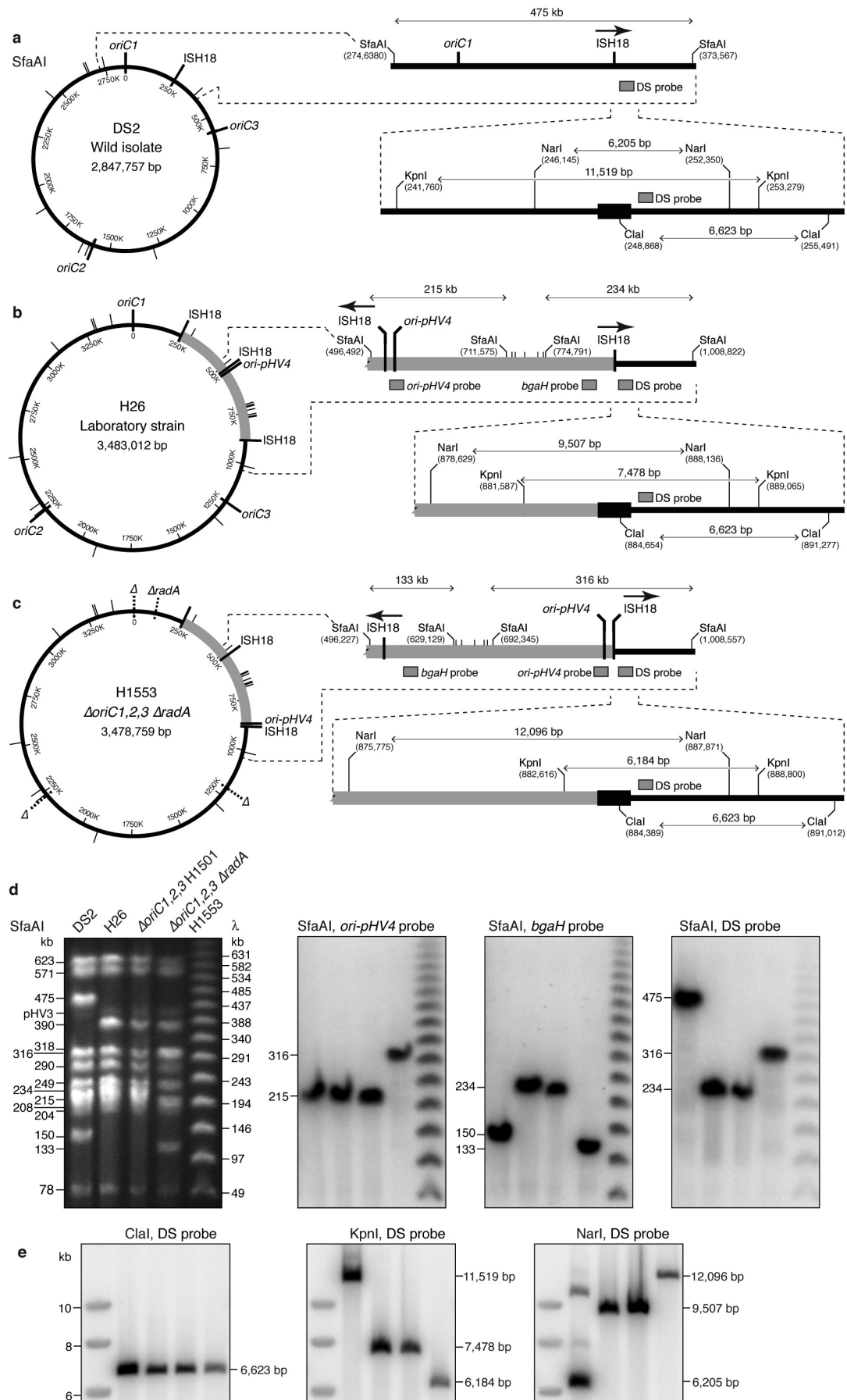
Extended Data Figure 5 | Identifying chromosomal rearrangement in *ΔoriC1,2,3 ΔradA* **strain H1553. a**, Map of SfaAI restriction sites on the main chromosome of wild isolate DS2. The region around ISH18 HVO_0278 is shown with additional restriction sites and the probe. **b**, Map of SfaAI restriction sites on the main chromosome of laboratory strain H26. The region downstream of integrated pHV4 is shown with the same restriction sites as in Extended Data Fig. 5a, and two extra probes (*ori-pHV4* and *bgaH*) that hybridize to pHV4. **c**, Map of SfaAI restriction sites on the main chromosome of *ΔoriC1,2,3 ΔradA* strain H1553. The region downstream of integrated pHV4 is shown as in Extended Data Fig. 5b. H1553 has undergone a chromosomal rearrangement involving part of pHV4 between ISH18 HVO_A0014 and ISH18 HVO_0278. These ISH18 elements are identical in sequence but in an inverted orientation (bold arrows); recombination between them results in inversion of the intervening sequence. **d**, Restriction fragment length polymorphisms in H26 and H1553. Genomic DNA from wild isolate DS2, laboratory strain H26, *ΔoriC1,2,3* strain H1501 and *ΔoriC1,2,3 ΔradA* strain H1553 was digested with SfaAI and shown on a pulsed-field gel. Southern blots were probed with the *ori-pHV4* origin, *bgaH* gene (located on pHV4 (ref. 21)) and sequences downstream (DS) of ISH18 element HVO_0278. **e**, Confirmation of restriction fragment length polymorphisms by ClaI, KpnI and NarI digests, probed with sequences downstream (DS) of ISH18 element HVO_0278; see also Extended Data Fig. 4.
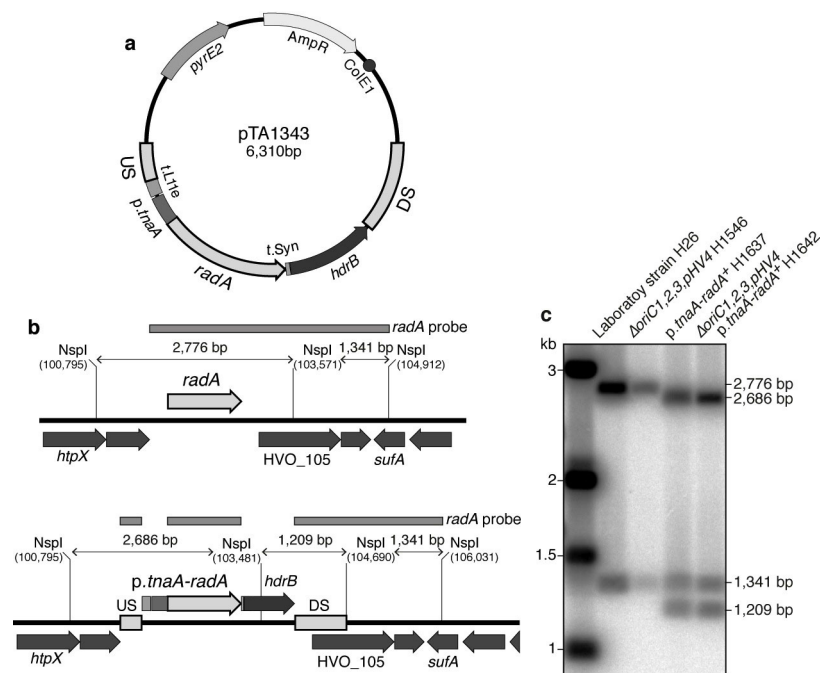
Extended Data Figure 6 | Generating tryptophan-inducible *radA* strains.
a, Map of p.*tnaA-radA*[+] gene replacement plasmid pTA1343. **b**, Map of region around *radA*, showing NspI restriction sites and the probe used to determine replacement of the native *radA* gene with the tryptophan-inducible *radA* allele. **c**, Confirmation of *radA* replacement by p.*tnaA-radA*[+]::*hdrB*[+]. Genomic DNA from laboratory strain H26, *ΔoriC1,2,3,pHV4* strain H1546, p.*tnaA-radA*[+] strain H1637 and *ΔoriC1,2,3,pHV4* p.*tnaA-radA*[+] strain H1642 was digested with NspI and probed with the *radA* region. See Methods for further details.

Extended Data Table 1 | SNPs identified in deep sequencing data

Strain	Chromosome				pHV1	pHV4	
	bp 31720 HVO_0032 Hypothetical	bp 985862 HVO_1080 Hypothetical	bp 1142038 HVO_1254 Hypothetical	bp 2410564 HVO_2547 *rpl32e*	bp 62044 HVO_C0064 Transposase	bp 626377 HVO_A0627 Hypothetical	bp 631510 HVO_A0634 Peptidase
Reference[2]	C	G	G	A	C	C	G
DS2 wild-isolate	C	G	G	G	C	C	G
H26 laboratory strain	T	T	G	G	C	C	G
(H53)	t	t	g	g	c	C	A
H1269 Δ*oriC1*	T	T	A	G	C	C	A
H1267 Δ*oriC2*	T	T	G	G	A	C	A
H1371 Δ*oriC3*	T	T	G	G	C	C	A
H1539 Δ*ori-pHV4*	T	T	G	G	C	C	A
(H1293 Δ*oriC1::trpA+* Δ*oriC2*)	t	t	g	g	c	C	A
(H1340 Δ*oriC1,2*)	t	t	g	g	c	A	A
H1374 Δ*oriC1,2,3*	T	T	G	G	C	A	A
H1546 Δ*oriC1,2,3,pHV4*	T	T	G	G	C	A	A
Amino acid changes	(Pro–>Pro)	Ser–>Tyr	Gly–>Ser	Leu–>Pro	(Arg–>Arg)	Gly–>Val	Val–>Ile

Locations of SNPs that differ from the published sequence are shown alongside predicted mutations. Strains in parentheses were subject to low coverage sequencing (bases in lower case were imputed). See Methods for further details.

Extended Data Table 2 | *H. volcanii* strains

Strain	Genotype	Reference/derivation*
DS2	Wild-type	36
DS70	ΔpHV2	37
H26	ΔpyrE2	30
H53	ΔpyrE2 ΔtrpA	30
H54	ΔpyrE2 bgaHa	21
H98	ΔpyrE2 ΔhdrB	30
H99	ΔpyrE2 ΔtrpA ΔhdrB	30
H112	ΔpyrE2 ΔradA	4
H230	ΔpyrE2 ΔtrpA Δori-pHV1::trpA⁺	4
H300	ΔpyrE2 ΔtrpA Δori-pHV1::trpA⁺ ΔhdrB	H230 pTA155
H1023	ΔpyrE2 ΔtrpA Δori-pHV1::trpA⁺ ΔhdrB ΔoriC1::hdrB⁺	H300 pTA946
H1267	ΔpyrE2 ΔtrpA ΔoriC2	H53 pTA1209
H1268	ΔpyrE2 ΔtrpA ΔoriC2	H53 pTA1209
H1269	ΔpyrE2 ΔtrpA ΔoriC1	H53 pTA1208
H1293	ΔpyrE2 ΔtrpA ΔoriC1::trpA⁺ ΔoriC2	H1268 pTA532
H1340	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2	H1293 pTA1208
H1371	ΔpyrE2 ΔtrpA ΔoriC3::trpA⁺	H53 pTA1249
H1372	ΔpyrE2 ΔtrpA ΔoriC2 ΔoriC3::trpA⁺	H1267 pTA1249
H1373	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC3::trpA⁺	H1269 pTA1249
H1374	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3::trpA⁺	H1340 pTA1249
H1458	ΔpyrE2 ΔtrpA ΔoriC3	H1371 pTA1248
H1460	ΔpyrE2 ΔtrpA ΔoriC2 ΔoriC3	H1372 pTA1248
H1462	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC3	H1373 pTA1248
H1464	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3	H1374 pTA1248
H1495	ΔpyrE2 ΔtrpA ΔoriC1 ΔhdrB	H1269 pTA155
H1496	ΔpyrE2 ΔtrpA ΔoriC2 ΔhdrB	H1267 pTA155
H1497	ΔpyrE2 ΔtrpA ΔoriC3 ΔhdrB	H1458 pTA155
H1501	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3 ΔhdrB	H1464 pTA155
H1539	ΔpyrE2 ΔtrpA Δori-pHV4::trpA⁺	H53 pTA1331
H1540	ΔpyrE2 ΔtrpA ΔoriC1 Δori-pHV4::trpA⁺	H1269 pTA1331
H1541	ΔpyrE2 ΔtrpA ΔoriC2 Δori-pHV4::trpA⁺	H1267 pTA1331
H1542	ΔpyrE2 ΔtrpA ΔoriC3 Δori-pHV4::trpA⁺	H1458 pTA1331
H1543	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 Δori-pHV4::trpA⁺	H1340 pTA1331
H1544	ΔpyrE2 ΔtrpA ΔoriC2 ΔoriC3 Δori-pHV4::trpA⁺	H1460 pTA1331
H1545	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC3 Δori-pHV4::trpA⁺	H1462 pTA1331
H1546	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3 Δori-pHV4::trpA⁺	H1464 pTA1331
H1547	ΔpyrE2 ΔtrpA ΔhdrB ΔradA::trpA⁺	H99 pTA324†
H1548	ΔpyrE2 ΔtrpA ΔoriC1 ΔhdrB ΔradA::trpA⁺	H1495 pTA324†
H1549	ΔpyrE2 ΔtrpA ΔoriC2 ΔhdrB ΔradA::trpA⁺	H1496 pTA324†
H1550	ΔpyrE2 ΔtrpA ΔoriC3 ΔhdrB ΔradA::trpA⁺	H1497 pTA324†
H1553	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3 ΔhdrB ΔradA::trpA⁺	H1501 pTA324†
H1593	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3 ΔhdrB Δori-pHV4	H1501 pTA1329
H1596	ΔpyrE2 ΔtrpA ΔoriC1 ΔoriC2 ΔoriC3 ΔhdrB Δori-pHV4 radA⁺::[ΔradA::trpA⁺ pyrE2⁺] { radA⁺ pyrE2⁺ hdrB⁺}	H1593 pTA324† No ΔradA obtained‡
H1608	ΔpyrE2 ΔoriC1 ΔoriC2 ΔoriC3 ΔhdrB Δori-pHV4	H1593 pTA49§
H1637	ΔpyrE2 ΔhdrB p.tnaA-radA⁺::hdrB⁺	H98 pTA1343
H1642	ΔpyrE2 ΔoriC1 ΔoriC2 ΔoriC3 ΔhdrB Δori-pHV4 p.tnaA-radA⁺::hdrB⁺	H1608 pTA1343

* Unless stated otherwise, source of strains is this study; parental strains and plasmids used in gene deletion are given for new strains.
† Plasmid pTA411 also used in deletion of *radA*[21].
‡ *radA* could not be deleted from H1596. Genes shown within square brackets are present on integrated *ΔradA* construct pTA324, genes shown within curly brackets are present on episomal *radA*⁺ complementation plasmid pTA411 (ref. 21).
§ Transformation to *trpA*⁺ using linear BstXI–BamHI fragment of pTA49 (ref. 30).

Extended Data Table 3 | Plasmids

Plasmid	Relevant properties	Source*
pTA49	pBluescript with *trpA* region	30
pTA128	pBluescript with *bgaH* gene from pHV4	21
pTA131	Integrative vector based on pBluescript II, with $pyrE2^+$ marker	30
pTA155	pTA131 with Δ*hdrB* construct	30
pTA187	pUC19 with $hdrB^+$ marker	30
pTA298	pUC19 with $trpA^+$ marker flanked by *Bam*HI sites	38
pTA324	pTA131 with Δ*radA::trpA*$^+$ construct	21
pTA411	pTA409 shuttle vector[21] with $radA^+$ gene, for complementation of Δ*radA*	21
pTA416	pBluescript with *oriC2* region	4
pTA441	pTA131 with *oriC1* region	4
pTA532	pTA441 with replacement of 470 bp *Eco*RI-*Bsp*EI *oriC1* origin fragment by 969 bp *Bam*HI *trpA*$^+$ from pTA298	
pTA612	pTA131 with *oriC2* region	4
pTA927	pTA230 shuttle vector[30] with t.L11e terminator, p.*tnaA* promoter and t.Syn terminator	33
pTA946	pTA441 with replacement of 470 bp *Eco*RI-*Bsp*EI *oriC1* origin fragment by 716 bp *Bam*HI-*Xba*I *hdrB*$^+$ fragment from pTA187	
pTA1100	pTA131 with 4.67 kb *Aci*I fragment from H1023 containing *oriC3* region, inserted at *Cla*I site	
pTA1208	pTA441 with deletion of 470 bp *Eco*RI-*Bsp*EI *oriC1* origin fragment	
pTA1209	pTA131 with 5.47 kb *Nhe*I–*Eco*RI *oriC2* region fragment from pTA416 inserted at *Xba*I and *Eco*RI sites, and deletion of 856 bp *Rsr*II *oriC2* origin fragment	
pTA1236	pBluescript with 7.48 kb *Kpn*I fragment from H26 containing region downstream of pHV4 integration into chromosome, inserted at *Kpn*I site	
pTA1238	pBluescript with 3.65 kb *Nar*I fragment from H26 containing region upstream of pHV4 integration into chromosome, inserted at *Cla*I site	
pTA1248	pTA131 with 3.16 kb *Aci*I-*Stu*I *oriC3* region fragment from pTA1100 inserted at *Cla*I and *Xho*I sites, and deletion of 550 bp *Rsr*II *oriC3* origin fragment	
pTA1249	As pTA1248, but replacement of 550 bp *Rsr*II *oriC3* origin fragment by 969 bp *Bam*HI *trpA*$^+$ from pTA298	
pTA1329	pTA131 with 4.54 kb Δ*ori-pHV4* construct, consisting of 2.0 kb upstream *Kpn*I–*Bam*HI and 2.54 kb downstream *Bam*HI–*Xba*I PCR fragments ligated at internal *Bam*HI sites, and inserted at *Kpn*I and *Xba*I sites	
pTA1331	pTA1329 with insertion of 969 bp *Bam*HI *trpA*$^+$ fragment from pTA298 at internal *Bam*HI site	
pTA1343	pTA131 with 299 bp upstream *Kpn*I–*Bam*HI and 721 bp downstream *Bam*HI–*Eco*RI regions of *radA*, flanking a 2.16 kb *Bgl*II p.*tnaA-radA*$^+$*::hdrB*$^+$ construct, generated by inserting 1.04 kb *Nde*I–*Bam*HI *radA* PCR fragment downstream of p.*tnaA* promoter in pTA927, and 716 bp *Bam*HI-*Xba*I *hdrB*$^+$ fragment of pTA187	
pCN27	pTA131 with *ori-pHV4* origin	4
pSJS1140	pUC118 with *radA* region	20

* Unless stated otherwise, source of plasmids is this study.

Extended Data Table 4 | Oligonucleotides and probes

a

Primer	Sequence (5'-3')	Relevant properties*
RFBF	CCACGATGCCTTCGCACCTG	Forward PCR primer to generate DS probe downstream of ISH18 element HVO_0278
RFBR	CCACGATGCCTTCGCACCTG	Reverse PCR primer to generate DS probe downstream of ISH18 element HVO_0278
RFB5F	CGGGTCTTTGGTTAGTCAGGG	Forward PCR primer to generate US probe upstream of ISH18 element HVO_0278
RFB3R	CGGGGGATGAGTGGGATAGG	Reverse PCR primer to generate US probe upstream of ISH18 element HVO_0278
dOri_pHv42_UF	TTCA<u>GGTACC</u>TAACGTGGAACTACGG	Forward PCR primer to amplify upstream region of Δori-pHV4 construct, KpnI site
dOri_pHv42_UR	ACTGC<u>GGATCC</u>AGTGGTGTTGTAGGG	Reverse PCR primer to amplify upstream region of Δori-pHV4 construct, BamHI site
dOri_pHv42_DF	GAACG<u>GGATCC</u>GCGGACACTCCGGACGC	Forward PCR primer to amplify downstream region of Δori-pHV4 construct, BamHI site
dOri_pHv42_DR	GACAT<u>TCTAGA</u>CCGACTCGACCGGCTCG	Reverse PCR primer to amplify downstream region of Δori-pHV4 construct, XbaI site
pBSR3	ACCCCAGGCTTTACACTTTATGC	Forward PCR primer to amplify upstream flanking region of radA from pSJS1140[20], internal KpnI site used
dradAUSR	CTTCTG<u>GGATCC</u>CCAGTCGTTCCGCC	Reverse PCR primer to amplify upstream flanking region of radA, BamHI site
dradADSF	GCCGT<u>GGATCC</u>GTCGGCCGCTCAATCAC	Forward PCR primer to amplify downstream flanking region of radA, BamHI site
radAR2	ACCAACAGGTCGTAGTCCACCTCC	Reverse PCR primer to amplify downstream flanking region of radA, internal EcoRI site used
radANdeF	GAACGACTG<u>CATATG</u>GCAGAAGACG	Forward PCR primer to amplify radA gene, NdeI site
radABamR	CCGAC<u>GGATCC</u>ACGGCTTACTCGG	Reverse PCR primer to amplify radA gene, BamHI site

b

Probe	Usage	Location	Source
oriC1	Fig. 2a	oriC1 origin	470 bp EcoRI–BspEI fragment of pTA441[4]
oriC2	Fig. 2a	oriC2 origin	856 bp RsrII fragment of pTA416[4]
oriC3	Fig. 2a	oriC3 origin	550 bp RsrII fragment of pTA1100
ori-pHV4	Fig. 2a, Extended Data Fig. 5d	ori-pHV4 origin	598 bp BamHI–EcoRI fragment of pCN27[4]
US	Extended Data Fig. 4c	Main chromosome, upstream of ISH18 element HVO_0278	804 bp PCR using RFB5F and RFB3R
DS	Extended Data Figs. 4c, 5d and 5e	Main chromosome, downstream of ISH18 element HVO_0278	902 bp PCR using RFBF and RFBR
bgaH	Extended Data Fig. 5d	pHV4, bgaH gene	2581 bp HindIII–BsrGI fragment of pTA128[21]
radA	Extended Data Fig. 6c	radA region	3.41 kb KpnI–SphI fragment of pSJS1140[20]

(a) Oligonucleotides
* Restriction sites used in cloning are underlined.
(b) Probes

LETTER

K–Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions

Jonathan M. Ostrem[1]*, Ulf Peters[1]*, Martin L. Sos[1], James A. Wells[2] & Kevan M. Shokat[1]

Somatic mutations in the small GTPase K-Ras are the most common activating lesions found in human cancer, and are generally associated with poor response to standard therapies[1–3]. Efforts to target this oncogene directly have faced difficulties owing to its picomolar affinity for GTP/GDP[4] and the absence of known allosteric regulatory sites. Oncogenic mutations result in functional activation of Ras family proteins by impairing GTP hydrolysis[5,6]. With diminished regulation by GTPase activity, the nucleotide state of Ras becomes more dependent on relative nucleotide affinity and concentration. This gives GTP an advantage over GDP[7] and increases the proportion of active GTP-bound Ras. Here we report the development of small molecules that irreversibly bind to a common oncogenic mutant, K-Ras(G12C). These compounds rely on the mutant cysteine for binding and therefore do not affect the wild-type protein. Crystallographic studies reveal the formation of a new pocket that is not apparent in previous structures of Ras, beneath the effector binding switch-II region. Binding of these inhibitors to K-Ras(G12C) disrupts both switch-I and switch-II, subverting the native nucleotide preference to favour GDP over GTP and impairing binding to Raf. Our data provide structure-based validation of a new allosteric regulatory site on Ras that is targetable in a mutant-specific manner.

To target K-Ras(G12C) we took advantage of the unique nucleophilicity of cysteine thiols by exploring cysteine-reactive small molecules. This strategy has the added advantage of allowing selectivity for the mutant over wild-type K-Ras. Notably, the mutant Cys 12 sits in close proximity to both the nucleotide pocket and the switch regions involved in effector interactions (Fig. 1a). To identify a chemical starting point, we used a disulphide-fragment-based screening approach called tethering[8]. We screened a library of 480 tethering compounds against K-Ras(G12C) in the GDP state using intact protein mass spectrometry[9,10] (see Methods and Extended Data Table 1). Fragments 6H05 (94 ± 1% (mean ± s.d.)) and 2E07 (84.6 ± 0.3%) gave the greatest degree of modification (Fig. 1b, c). Reaction with wild-type K-Ras, which contains three native cysteine residues, was not detected. Conversely, both compounds modify the oncogenic G12C mutant of the highly homologous protein H-Ras[11,12] (Fig. 1b). Binding was not diminished by 1 mM GDP in the presence of EDTA, suggesting that the compounds bind in an allosteric site not overlapping with GDP. Pre-loading of K-Ras with GTP significantly impairs modification by both compounds, indicating incompatibility between compound binding and the active conformation of Ras.

We chose to pursue the top fragment, 6H05, by investigating structure–activity relationships for several analogues[8] (Fig. 1c, see Methods). Some changes such as replacing the thioether with a methylene group reduced binding (1, relative potency < 0.1). However, other changes such as modification of the tethering linker enhanced binding (6, relative potency 4.2). Having identified a tractable chemotype, we pursued a co-crystal structure to enable structure-based design. To facilitate uniform labelling on Cys 12 we used a K-Ras construct lacking other cysteines,
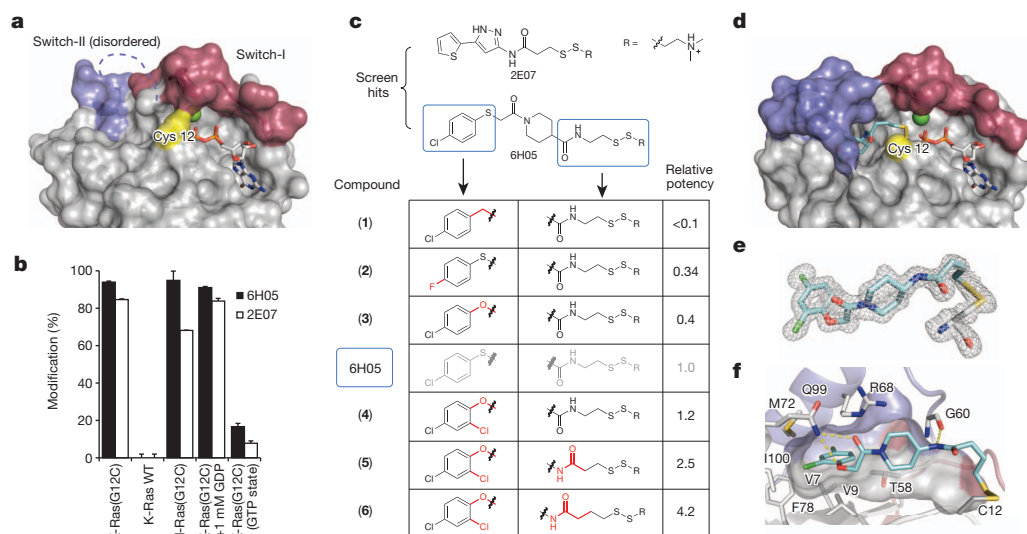
Figure 1 | Tethering compounds selectively bind to oncogenic K–Ras(G12C). a, Crystal structure of K-Ras(G12C) GDP shows Cys 12 (yellow), switch-I (red) and switch-II (blue). Switch-II is partially disordered. **b**, Percentage modification by compounds 6H05 and 2E07 ($n = 3$, error bars denote s.d.). **c**, 6H05 analogue structure–activity relationship. Relative potency = (fragment DR_{50})/(6H05 DR_{50}), in which DR_{50} denotes the dose ratio resulting in 50% modification; see Methods. **d**, Co-crystal structure of **6** (cyan) and K-Ras(G12C) with GDP (grey) and Ca^{2+} (green). **e**, $F_o - F_c$ omit map (grey mesh, 2.5σ) of **6** and Cys 12 from **d**. **f**, Surface representation of S-IIP around **6** showing hydrogen bonds (yellow lines). Indicated residues make hydrophobic contacts with **6**.

[1]Department of Cellular and Molecular Pharmacology, Howard Hughes Medical Institute, University of California, San Francisco, California 94158, USA. [2]Departments of Pharmaceutical Chemistry and Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, USA.
*These authors contributed equally to this work.

K-Ras(C51S/C80L/C118S) (known as Cys-light), which showed minimal effects on overall protein structure (G12C versus Cys-light, root mean squared deviation (r.m.s.d.) (Cα) = 0.33 Å). Using this construct, we obtained a 1.29 Å co-crystal structure of **6** bound to K-Ras(G12C) in the GDP state (Fig. 1d). Compound **6** does not bind in the nucleotide pocket but extends from Cys 12 into an adjacent pocket composed largely of switch-II. This fully formed pocket is not apparent in other published structures of Ras, although a groove is visible in some cases[13] (Extended Data Fig. 1b), and previous studies have suggested the presence of an allosteric site in this region[14]. We refer to the compound binding region as the switch-II pocket (S-IIP).

The S-IIP is located between the central β-sheet of Ras, and the α2- (switch-II) and α3-helices. Well-defined electron density shows the location of **6** deep within the S-IIP (432 Å² interface; Extended Data Table 2) and confirms the disulphide linkage between **6** and Cys 12 (Fig. 1e; $F_o - F_c$ at 2.5σ). The hydrophobic dichlorophenyl group of **6** makes several hydrophobic contacts (Fig. 1f). Glu 99 and Gly 60 form direct hydrogen bonds to **6**. Whereas switch-II shows significant reordering to form the S-IIP, the conformation of switch-I is unchanged from the GDP-bound state. Structural analysis also suggested the presence of sub-pockets in S-IIP that might enable design of more potent inhibitors (Extended Data Fig. 2a, (o-) and (p-)).

Rather than continue with disulphide-based compounds we turned to carbon-based electrophiles, acrylamides and vinyl sulphonamides, which are still chemoselective yet provide irreversible cysteine bond formation. We synthesized nearly 100 analogues guided by iterative structural evaluation to yield substantial improvements in potency (Fig. 2a–c and Supplementary Table 1). Owing to the irreversible nature of binding, potency was assessed by time-dependent modification of the protein,

initially 200 μM compound for 24 h. A shift to 10 μM compound was necessary to differentiate optimized analogues (Extended Data Table 3 and Extended Data Fig. 3a), reaching the detection limit of the assay (4 μM K-Ras). Although vinyl sulphonamides generally performed better by these metrics, probably owing to higher reactivity[15], we obtained highly effective acrylamides as well (Fig. 2b).

To evaluate the off-target specificity of our most potent acrylamide **12**, we used intact protein mass spectrometry to monitor simultaneously for modification of K-Ras(G12C) and bovine serum albumin (BSA) (one free cysteine) in a single mixture. Treatment with **12** resulted in the modification of K-Ras(G12C) but not BSA (Extended Data Fig. 3b), although both react with Ellman's reagent (also known as DTNB) (Extended Data Fig. 3c). Optimized electrophiles show no detectable modification of wild-type K-Ras (Extended Data Fig. 3d). Fragments lacking the electrophile did not impair binding of compound **12** to K-Ras(G12C) (data not shown), suggesting limited binding to S-IIP in K-Ras proteins lacking the G12C mutation.

Overlaying multiple co-crystal structures revealed that the compounds follow a similar trajectory through the pocket and project functional groups into the (o-) and (p-) sub-pockets (Fig. 2c and Extended Data Fig. 2). Despite considerable variation at the terminal phenyl ring, the compounds satisfy similar hydrophobic interactions, supporting the crucial role of this region of the S-IIP (Fig. 2c and Extended Data Fig. 2).

In the active state of Ras, residues from switch-II entirely fill the S-IIP (Fig. 2d). Co-crystal structures with tethering compound **6** and the electrophile **8** exhibit displacement of switch-II relative to the active conformation (Fig. 2e, f). Comparison of these co-crystal structures revealed distinct effects on switch-I and switch-II. Tethering compound **6** induces a small displacement of switch-II with little effect
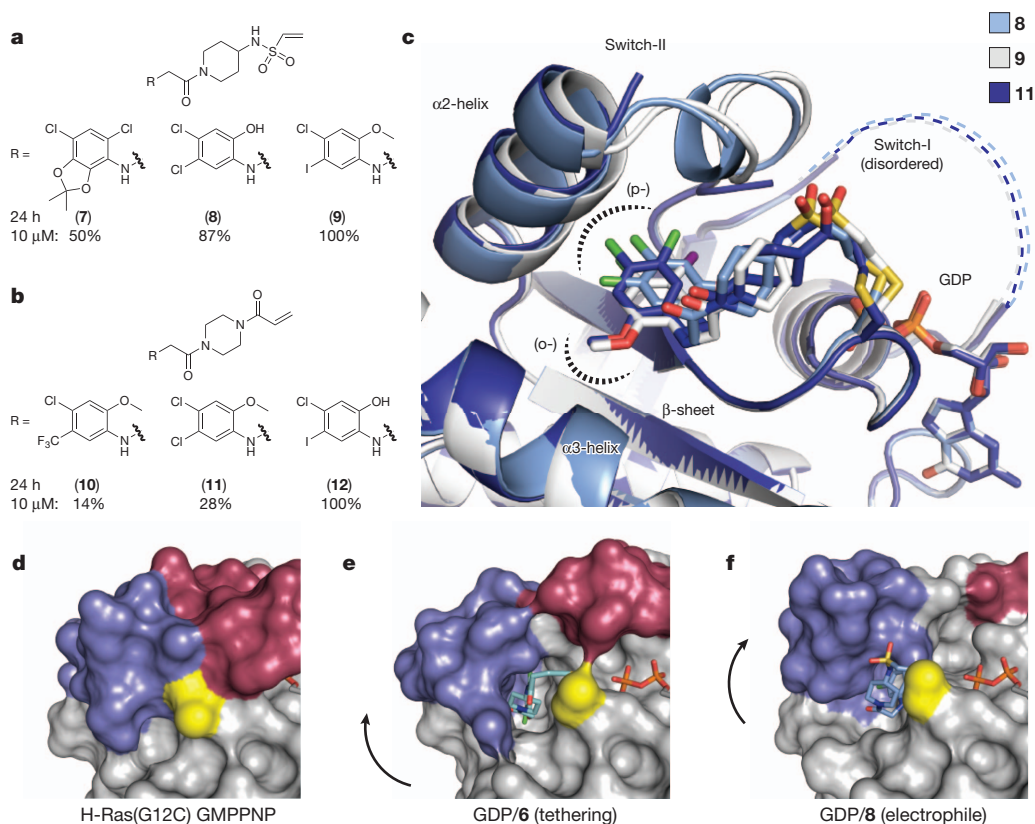
Figure 2 | **Electrophilic compounds bind to S-IIP of K-Ras(G12C) and disrupt switch-I and switch-II. a**, Subset of vinyl sulphonamide analogues. **b**, Subset of acrylamide analogues. Percentages in **a** and **b** represent adduct formation after 24 h with 10 μM compound. **c**, Overlay of co-crystal structures of 8, 9 and 11 with GDP-bound K-Ras(G12C). **d–f**, Binding of tethering compound **6** or electrophilic compound **8** to Cys 12 (yellow) of K-Ras GDP leads to displacement (arrows) of switch-II (blue) as compared to active Ras (H-Ras(G12C) GMPPNP; **d**). In the case of tethering compound **6** (**e**), switch-I (red) resembles the inactive GDP-bound conformation, however electrophile **8** (**f**) causes partial disordering of switch-I.
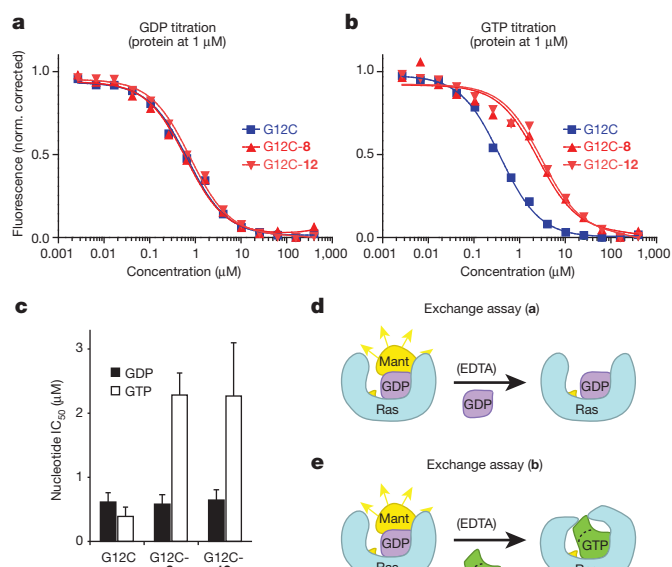
Figure 3 | Compound binding to S-IIP changes nucleotide preference of K-Ras from GTP to GDP. a, EDTA-mediated competition between mant-dGDP loaded on K-Ras(G12C) and free unlabelled GDP. The experiment was carried out with full-length K-Ras(G12C) alone (squares), or modified by **8** (upwards triangles) or **12** (downwards triangles) ($n = 3$). Data from a representative experiment is shown fitted to a sigmoidal curve for each protein. **b**, EDTA-mediated competition between bound mant-dGDP and free unlabelled GTP. **c**, Quantification of the GDP and GTP titrations in **a** and **b** ($n = 3$; error bars denote s.d.; IC_{50} obtained from sigmoidal fits). **d**, **e**, Schematic representation of experiments shown in **a** (**d**) and **b** (**e**).

on switch-I. By contrast, electrophile **8** induces a more pronounced displacement of switch-II that results in disordering of switch-I and a lack of density for the metal ion.

Proper metal coordination is crucial for tight nucleotide binding[16], with mutation of magnesium-coordinating residues Ser 17 or Asp 57 leading to a preference for GDP over GTP[17,18]. Many of our structures with carbon-based electrophiles show disordering of switch-I and a lack

of density for the metal ion (Extended Data Table 4). On the basis of these observations, we predicted that S-IIP binding compounds might differentially affect nucleotide affinities leading to a preference for GDP over GTP. To test this prediction, we carried out EDTA-catalysed off-exchange reactions with 2'-deoxy-3'-O-(N-methylanthraniloyl) (mant)-dGDP, while titrating unlabelled GDP or GTP (Fig. 3). In the absence of inhibitor, K-Ras(G12C) shows a slight preference for GTP (relative affinity 0.6 ± 0.2), as reported previously for H-Ras[7]. However, in the presence of either **8** or **12**, GTP affinity is significantly decreased relative to GDP (relative affinity 3.9 ± 0.6 (**8**) and 3.5 ± 0.8 (**12**), $P = 0.004$, t-test) (Fig. 3b, c). The catalysis of nucleotide exchange by EDTA suggests that although our compounds may subtly affect metal binding leading to changes in nucleotide affinity, Mg^{2+} is not precluded even when the S-IIP is occupied.

Structural analysis also predicts that the function of the exchange factor SOS would be compromised by compound binding to S-IIP[19]. Indeed, treatment of K-Ras(G12C) with either **8** or **12** blocks SOS-catalysed nucleotide exchange (Fig. 4a–e). As shown above, these compounds do not impair EDTA-catalysed GDP exchange.

In the active conformation of Ras, Gly 60 and Thr 35 make crucial contacts with the γ-phosphate[20]. Conservative mutations of Thr 35 (T35S) or Gly 60 (G60A) markedly impair effector binding[21,22]. Our compounds occupy the required position for Gly 60 in the active conformation and displace this residue to varying degrees, with larger distances correlating with disordering of switch-I (Extended Data Table 4 and Extended Data Fig. 4). This analysis led to the prediction that our compounds would disrupt the conformation of the GTP state of Ras and impair interactions with effectors such as Raf. We measured Ras–Raf association in two K-Ras(G12C)-mutant lung cancer cell lines, H1792 and H358, using co-immunoprecipitation. As predicted, treatment with **12** decreased the association of B-Raf and C-Raf with Ras (Fig. 4f). Although this effect is evident in both cell lines, it is most pronounced in H1792 cells, which express lower levels of Ras[23].

Despite limited compound potency, we speculated that the genotype-specificity might afford a therapeutic window for the targeted inhibition of K-Ras(G12C) in cellular models. Therefore, we compared the effects of **12** in a small collection of genetically annotated lung cancer cell lines. As expected, the group of cell lines containing G12C mutations (H1792, H358, H23 and Calu-1) showed decreased viability (Fig. 4g and Extended
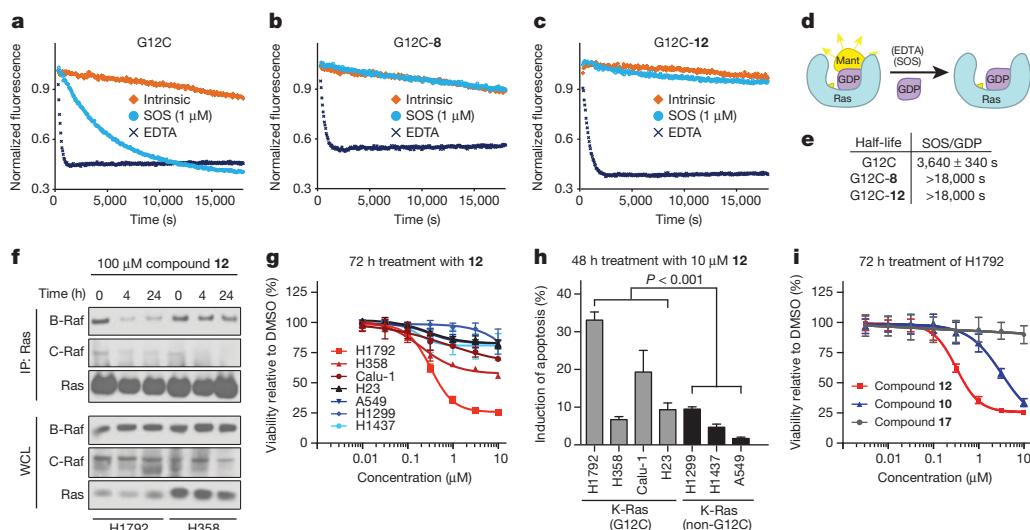
Figure 4 | Compounds block K-Ras(G12C) interactions, decrease viability and increase apoptosis of G12C-containing lung cancer cell lines. a–c, SOS-catalysed nucleotide exchange for full-length K-Ras(G12C) alone (**a**), or K-Ras(G12C) labelled with **8** (**b**) or **12** (**c**). **d**, Schematic representation of **a–c**. **e**, Half-life of exchange for **a–c** ($n = 3$ biological replicates, error bars denote s.d.). **f**, Co-immunoprecipitation (IP) of B-Raf and C-Raf with Ras from K-Ras(G12C) cell lines after treatment with compound **12** ($n = 3$ biological

replicates). WCL, whole cell lysate. **g**, Viability of K-Ras(G12C)-mutant cell lines (H1792, H358, Calu-1 and H23) and cell lines lacking this mutation (A549, H1299 and H1437) after treatment with **12** ($n = 3$ biological replicates, error bars denote s.e.m.). **h**, Induction of apoptosis after 48 h with 10 μM **12**. **i**, H1792 cell viability assays carried out as in **g**, with range of concentrations of **10**, **12** and **17** ($n = 3$ biological replicates, error bars denote s.e.m.).

Data Fig. 5; $P = 0.005$, t-test) and increased apoptosis (Fig. 4h; $P < 0.001$, t-test) relative to the group lacking this mutation (H1437, H1299 and A549) after treatment with **12**. The highly sensitive H1792 cells show low levels of K-Ras GTP (Extended Data Fig. 5b), consistent with preferential binding of our inhibitors to K-Ras GDP (Fig. 1b), and they are highly K-Ras dependent[24] (Extended Data Fig. 5c). Notably, both K-Ras-dependent (A549) and -independent (H1299) cell lines lacking G12C were insensitive to compound **12** (Fig. 4g, h and Extended Data Fig. 5). The half-maximum effective concentration (EC_{50}) for compound **12** in H1792 cells ($0.32 \pm 0.01\ \mu M$) is tenfold lower than that of compound **10** ($3.2 \pm 0.4\ \mu M$), consistent with *in vitro* K-Ras labelling efficiency (Fig. 2b, 100% versus 14% labelling at 24 h). Notably, the highly related electrophile-containing compound **17**, which does not modify K-Ras(G12C) *in vitro* (Extended Data Fig. 2c, 0% at 24 h), shows no effect on H1792 cell viability. Overall, our cellular data provide a proof-of-concept for the genotype-specific use of S-IIP binding compounds in K-Ras(G12C)-driven cancer.

Using a structure-guided approach to target the G12C mutant of K-Ras, we have identified a new allosteric pocket, S-IIP, in this protein, and we have used that pocket to develop irreversible, mutant-specific inhibitors of Ras function. The S-IIP is not visible in other structures of Ras, and thus it is probably highly dynamic when GDP is bound, until initial encounter with our compounds. Compound binding to S-IIP impairs Ras function through two distinct mechanisms. First, by shifting the relative nucleotide affinities of Ras to favour GDP over GTP, the compounds should lead to accumulation of Ras in the inactive state. Notably, the two most effective cellular GTPase inhibitors, the natural products brefeldin A and YM-254890, both bind to and stabilize a GDP-bound state of their respective GTPases[25,26]. So far, published small molecules that bind Ras have not shown this nucleotide preference[14,27–29]. Second, compounds occupying S-IIP diminish interactions with effectors and regulatory proteins. This should act to diminish signalling by K-Ras further. Despite the need for continued chemical optimization of our compounds for future assessment *in vivo*, initial evaluation of our compounds in lung cancer cell lines suggests allele-specific impairment of K-Ras function. On the basis of these data and our understanding of the biochemical mechanism of the inhibitors, we are confident that our findings can serve as the starting point for drug-discovery efforts targeting K-Ras(G12C) and eventually other alleles of K-Ras.

METHODS SUMMARY

Mass spectrometric analyses were carried out using Waters Acquity UPLC/ESI-TQD and Waters LCT-Premier LC/ESI-MS instruments. H23, H358, H1299, H1437, H1792, Calu-1 and A549 cells (ATCC) were cultured in DMEM with 10% FBS.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Slebos, R. J. C. *et al.* K-*ras* oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N. Engl. J. Med.* **323**, 561–565 (1990).
2. Pao, W. *et al.* KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med.* **2**, e17 (2005).
3. Lièvre, A. *et al.* KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res.* **66**, 3992–3995 (2006).
4. John, J. *et al.* Kinetics of interaction of nucleotides with nucleotide-free H-ras p21. *Biochemistry* **29**, 6058–6065 (1990).
5. Gibbs, J. B., Sigal, I. S., Poe, M. & Scolnick, E. M. Intrinsic GTPase activity distinguishes normal and oncogenic ras p21 molecules. *Proc. Natl Acad. Sci. USA* **81**, 5704–5708 (1984).
6. Trahey, M. & McCormick, F. A cytoplasmic protein stimulates normal N-ras p21 GTPase, but does not affect oncogenic mutants. *Science* **238**, 542–545 (1987).
7. Scherer, A. *et al.* Crystallization and preliminary X-ray analysis of the human c-H-ras-oncogene product p21 complexed with GTP analogues. *J. Mol. Biol.* **206**, 257–259 (1989).
8. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc. Natl Acad. Sci. USA* **97**, 9367–9372 (2000).
9. Burlingame, M. A., Tom, C. T. M. B. & Renslo, A. R. Simple one-pot synthesis of disulfide fragments for use in disulfide-exchange screening. *ACS Comb. Sci.* **13**, 205–208 (2011).
10. Sadowsky, J. D. *et al.* Turning a protein kinase on or off from a single allosteric site via disulfide trapping. *Proc. Natl Acad. Sci. USA* **108**, 6056–6061 (2011).
11. Forbes, S. A. *et al.* The catalogue of somatic mutations in cancer (COSMIC). *Curr. Protoc. Hum. Genet.* **57**, 10.11.1–10.11.26 (2008).
12. Bar-Sagi, D. A Ras by any other name. *Mol. Cell. Biol.* **21**, 1441–1443 (2001).
13. Milburn, M. V. *et al.* Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939–945 (1990).
14. Taveras, A. G. *et al.* Ras oncoprotein inhibitors: the discovery of potent, ras nucleotide exchange inhibitors and the structural determination of a drug–protein complex. *Bioorg. Med. Chem.* **5**, 125–133 (1997).
15. Naven, R. T., Kantesaria, S., Nadanaciva, S., Schroeter, T. & Leach, K. L. High throughput glutathione and Nrf2 assays to assess chemical and biological reactivity of cysteine-reactive compounds. *Toxicol. Rev.* **2**, 235–244 (2013).
16. John, J. *et al.* Kinetic and structural analysis of the Mg^{2+}-binding site of the guanine nucleotide-binding protein p21H-ras. *J. Biol. Chem.* **268**, 923–929 (1993).
17. Feig, L. A. & Cooper, G. M. Inhibition of NIH 3T3 cell proliferation by a mutant ras protein with preferential affinity for GDP. *Mol. Cell. Biol.* **8**, 3235–3243 (1988).
18. Farnsworth, C. L. & Feig, L. A. Dominant inhibitory mutations in the Mg^{2+}-binding site of Ras^H prevent its activation by GTP. *Mol. Cell. Biol.* **11**, 4822–4829 (1981).
19. Hall, B. E., Yang, S. S., Boriack-Sjodin, P. A., Kuriyan, J. & Bar-Sagi, D. Structure-based mutagenesis reveals distinct functions for Ras switch 1 and switch 2 in Sos-catalyzed guanine nucleotide exchange. *J. Biol. Chem.* **276**, 27629–27637 (2001).
20. Pai, E. F. *et al.* Structure of the guanine-nucleotide-binding domain of the Ha-*ras* oncogene product p21 in the triphosphate conformation. *Nature* **341**, 209–214 (1989).
21. Sung, Y.-J., Carter, M., Zhong, J.-M. & Hwang, Y.-W. Mutagenesis of the H-ras p21 at glycine-60 residue disrupts GTP-induced conformational change. *Biochemistry* **34**, 3470–3477 (1995).
22. Hwang, M.-C. C., Sung, Y.-J. & Hwang, Y.-W. The differential effects of the Gly-60 to Ala mutation on the interaction of H-Ras p21 with different downstream targets. *J. Biol. Chem.* **271**, 8196–8202 (1996).
23. Sunaga, N. *et al.* Knockdown of oncogenic KRAS in non-small cell lung cancers suppresses tumor growth and sensitizes tumor cells to targeted therapy. *Mol. Cancer Ther.* **10**, 336–346 (2011).
24. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic *KRAS*-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
25. Peyroche, A. *et al.* Brefeldin A acts to stabilize an abortive ARF–GDP–Sec7 domain protein complex. *Mol. Cell* **3**, 275–285 (1999).
26. Nishimura, A. *et al.* Structural basis for the specific inhibition of heterotrimeric G_q protein by a small molecule. *Proc. Natl Acad. Sci. USA* **107**, 13666–13671 (2010).
27. Maurer, T. *et al.* Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc. Natl Acad. Sci. USA* **109**, 5299–5304 (2012).
28. Sun, Q. *et al.* Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angew. Chem.* **124**, 6244–6247 (2012).
29. Shima, F. *et al.* In silico discovery of small-molecule Ras inhibitors that display antitumor activity by blocking the Ras-effector interaction. *Proc. Natl Acad. Sci. USA* **110**, 8182–8187 (2013).

METHODS

Protein expression and purification. Hexahistidine-tagged recombinant human K-Ras (isoform 2, residues 1–169, based on construct used for PDB accession 3GFT) was transformed into *Escherichia coli* (BL21 (DE3)). After the bacterial growth to an attenuance (*D*) at 600 nm of 0.4–0.6 in Terrific Broth containing 30 mg l^{-1} kanamycin at 37 °C, induction was carried out at 18 °C using 0.5 mM isopropyl-β-D-thiogalactoside (IPTG), and growth was continued at 18 °C for about 18 h. The bacteria were collected by centrifugation, and the obtained pellet either stored at −80 °C or used freshly for the subsequent steps.

The pellet was resuspended in lysis buffer (500 mM NaCl, 20 mM Tris, pH 8.0 and 5 mM imidazole) containing protease inhibitor cocktail (Roche complete EDTA free), the bacteria were lysed by microfluidizer, 2 mM β-mercaptoethanol (βME) (final) was added and cell debris was removed by ultracentrifugation. The supernatant was incubated for 1 h with Co-affinity beads (Clontech, ~2 ml bed volume per 1 l initial culture), the loaded beads were then washed with lysis buffer containing 2 mM βME and the protein was eluted with buffer containing 125–250 mM imidazole. The hexahistidine tag was then cleaved using hexahistidine-tagged TEV-protease (1 mg recombinant TEV per 25 mg crude K-Ras, 1 mg GDP added per 20 mg crude K-Ras) while dialysing against a buffer containing 300 mM NaCl, 20 mM Tris, pH 8.0, 5 mM imidazole, 1 mM dithiothreitol (DTT) and 0.5 mM EDTA. The cleaved protein was then diluted fivefold with low-salt buffer (50 mM NaCl, 20 mM Tris, pH 8.0), incubated with Ni-agarose beads (Qiagen) to remove uncleaved protein and protease, and 5 mM MgCl$_2$ and GDP was added to load the metal and nucleotide site of K-Ras fully.

The crude protein was then purified by ion-exchange chromatography (HiTrap Q HP column, salt gradient from 50 to 500 mM NaCl) to give the partially purified protein, commonly in the following buffer (~230 mM NaCl, 20 mM Tris, pH 8.0, small amounts of GDP). At this point the protein was either fully labelled with the desired compound (incubation overnight with an excess of compound at 4 °C, labelling checked by mass spectrometry analysis), frozen down and stored at −80 °C, or used for further purification.

The last purification step for the labelled or unlabelled protein was gel filtration using either a Superdex 75 or 200 column (10/300 GL) with the following buffer: 20 mM HEPES, pH 7.5, 150 mM NaCl and 1 mM DTT (for the unlabelled proteins). The freshly prepared and purified protein was then concentrated to 5–20 mg ml^{-1} and used for the X-ray crystallography trays.

Sequences for the different K-Ras constructs were generally codon-optimized and synthesized by DNA2.0 using the pJexpress411 vector. For the X-ray structures of compound-labelled K-Ras(G12C), a cysteine-light mutant was used (K-Ras (G12C/C51S/C80L/C118S)) to enable more uniformly labelled species.

Purification and labelling of full-length forms of the protein as well as H-Ras was carried out analogously. Nucleotide exchange for crystallographic samples was carried out following published procedures[4,27,30].

Tethering screen. Untagged recombinant K-Ras(G12C) (1–169) at 1 μM was reacted with 100 μM fragment and 100 μM βME in 20 mM HEPES, pH 7.5, 150 mM NaCl and 10 mM EDTA for 1 h at ambient temperature. The extent of modification was assessed by electrospray mass spectrometry using a Waters LCT-Premier LC/ESI-MS. By setting a threshold of ≥60% modification, we achieved a hit rate of 1.9%.

Determination of DR$_{50}$ and relative potency of fragments. The DR$_{50}$ is determined by titrating fragment while maintaining a constant βME concentration, in this case 200 μM βME. DR$_{50}$ = [βME]/[fragment], at which 50% modification is achieved by total protein mass spectrometry[10]. Relative potency is reported as: fragment DR$_{50}$/6H05 DR$_{50}$.

Mass spectrometric screen for extent of irreversible labelling. Untagged recombinant K-Ras(G12C) (1–169) at 4 μM was reacted with inhibitors at 200 μM or 10 μM (2% (v/v) dimethylsulphoxide (DMSO) final) in 20 mM HEPES, pH 7.5, 150 mM NaCl and 1 mM EDTA. After 24 h, 10-μl aliquots were removed and the reactions were stopped by the addition of 1 μl 2% (v/v) formic acid. For the BSA specificity experiment, 16 μM BSA was included in the mixture and the reaction was analysed at 6 h, without acid treatment. A similar mixture of K-Ras(G12C) and BSA was treated instead with 200 μM Ellman's reagent and labelling was analysed after 5 min. In all cases, the extent of modification was assessed by electrospray mass spectrometry using a Waters Acquity UPLC/ESI-TQD with a 2.1 × 50 mm Acquity UPLC BEH300 C4 column.

Plate-based assay to determine relative affinity of K-Ras for GDP or GTP. The corresponding recombinantly expressed, full-length K-Ras protein (G12C mutant or G12C mutant labelled fully with either compound **8** or **12**) at about 10 μM concentration was incubated with 200 μM mant-dGDP (Jena Biosciences) in the presence of 2.5 M EDTA. After 1 h at room temperature, MgCl$_2$ to a final concentration of 10 mM was added. The protein was then run through a NAP-5 column to remove free nucleotide. The concentration of the obtained protein was determined by Bradford assay and the protein was then used in the described plate-based assay.

For the assay, 10 μl of the prepared protein in reaction buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 1 mM DTT and 1 mM MgCl$_2$) was added to a well of a low volume black bottom plate (Corning, 3676). The fluorescence intensity was measured on a spectramax M5 plate reader (Molecular Devices, 360 nm excitation, 440 nm emission) to provide a value used in later normalization. Then, 5 μl of an EDTA solution with the indicated nucleotide (GDP or GTP) was added to each well and the reaction mix was allowed to equilibrate for 2 h at room temperature. Measurement of the fluorescent intensity at this time provided the end point. Samples were measured in duplicates for each experiment. In the final mix the concentrations were the following: protein (1 μM), EDTA (5 mM) and nucleotide (as indicated, titrated in 2.5-fold dilution series, 15 points). Curves show results from one representative experiment, the column graph shows the averaged data from three experiments, with errors representing s.d. For the determination of IC$_{50}$, a sigmoidal curve fit was used for each nucleotide (Prism software).

Nucleotide exchange rates for compound-bound K-Ras(G12C). The experiment was carried out similarly as described for the nucleotide affinity assay using the same plate set-up and plate reader. In brief, the respective full-length proteins were loaded with mant-dGDP (see above). For the assay, 10 μl of the prepared protein (1 μM final) in reaction buffer was added to the wells. To start the reaction, 5 μl of SOS (1 μM final), EDTA (5 mM final), or buffer was added and the fluorescence monitored for 5 h at 90-s intervals. Half-lives were determined using Prism software (single-exponential decay fit).

Cell culture. H23, H358, H1299, H1437, H1792, Calu-1 and A549 (ATCC) were cultured in DMEM with 10% FBS.

Viability assays. Cells were plated in 96-well plates at 2,000 cells per well in 90 μl DMEM with 10% FBS and allowed to attach for 24 h. Cells were treated by the addition of 10 μl 100 μM compound (or half-log dilutions thereof) or vehicle (0.1% DMSO final). After 72 h, media was exchanged and plates were analysed using CellTiter-Glo Luminescent Cell Viability Assay (Promega).

Apoptosis assays. The Annexin V-FITC Apoptosis Detection Kit I (BD Biosciences) was used to detect apoptotic cells. Cell lines were plated in 6-well plates at ~50% confluence, and after 24 h cells were treated with the given compound for 48 h. Subsequently, cells were washed with PBS, trypsinized and resuspended in 150 μl annexin-V binding buffer. Finally, cells were stained with annexin-V-FITC and propidium iodide and incubated in the dark before analysis on a FACS LSRII Flow Cytometer (Beckman Coulter). Data were collected using FACSDiva analysis software (Beckman Coulter). At least 10,000 cells were measured per individual sample. Results are given as Δannexin-V/propidium-iodide unstained cells of untreated (DMSO) versus treated samples.

siRNA knockdown. Cells were plated either in 96-well plates or 6-well plates 24 h before transfection with short interfering RNA (siRNA). The siRNA constructs were diluted (10–20 nM) in RNAiMax-Lipofectamine (Life Technologies) containing OPTI-MEM media, and after 20 min of incubation the mixture was added drop-wise to the cells. After 72 h of incubation, cells were either lysed for immunoblotting experiments (K-Ras antibody, Sigma 3B10-2F2; actin antibody, Cell Signaling Technology 4970) or subjected to CellTiterGlo assays (Promega) for proliferation analysis. The *KRAS* siRNA guide and passenger strand sequences used were as follows: guide, 5′-ACUGUACUCCUCUUGACCUGCU-3′; passenger, 5′-CAGGUCAAGAGGAGUACAGUUA-3′.

Immunoprecipitation. Cells were treated at given conditions in 10-cm plates and lysed in phosphatase and protease inhibitor containing lysis buffer. Lysates were mixed with 4 μl primary Ras antibody (Abcam, EPR3255), and incubated with rocking overnight at 4 °C. Protein G agarose beads were added to the mixture, and after 3 h beads were pelleted, washed with lysis buffer, and after resuspension in loading buffer and heating at 95 °C for 5 min were analysed by SDS–PAGE followed by immunoblot. B-Raf (Santa Cruz, F-7) and C-Raf (Cell Signaling Technology, 9422) antibodies were used to detect the individual proteins.

30. Ahmadian, M. R. *et al.* Guanosine triphosphatase stimulation of oncogenic Ras mutants. *Proc. Natl Acad. Sci. USA* **96**, 7065–7070 (1999).

a G12C-**6**

b Ref. 13 with **6** (overlayed)

c Ref. 27 with **6** (overlayed)
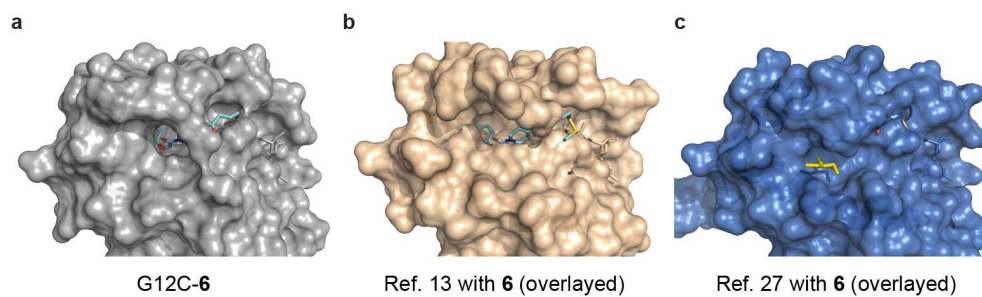
Extended Data Figure 1 | Comparison of co-crystal structure of 6 with K-Ras(G12C) to known structures of Ras. a, Compound **6** (cyan) bound in the S-IIP of K-Ras(G12C). **b**, Compound **6** (aligned and overlayed) with GDP-bound wild-type H-Ras showing groove near S-IIP (PDB accession 4Q21)[13]. **c**, Clash of compound **6** (aligned and overlayed) with GTPγS-bound K-Ras(G12D), which shows glycerol molecule adjacent to S-IIP (PDB accession 4DSO)[27].
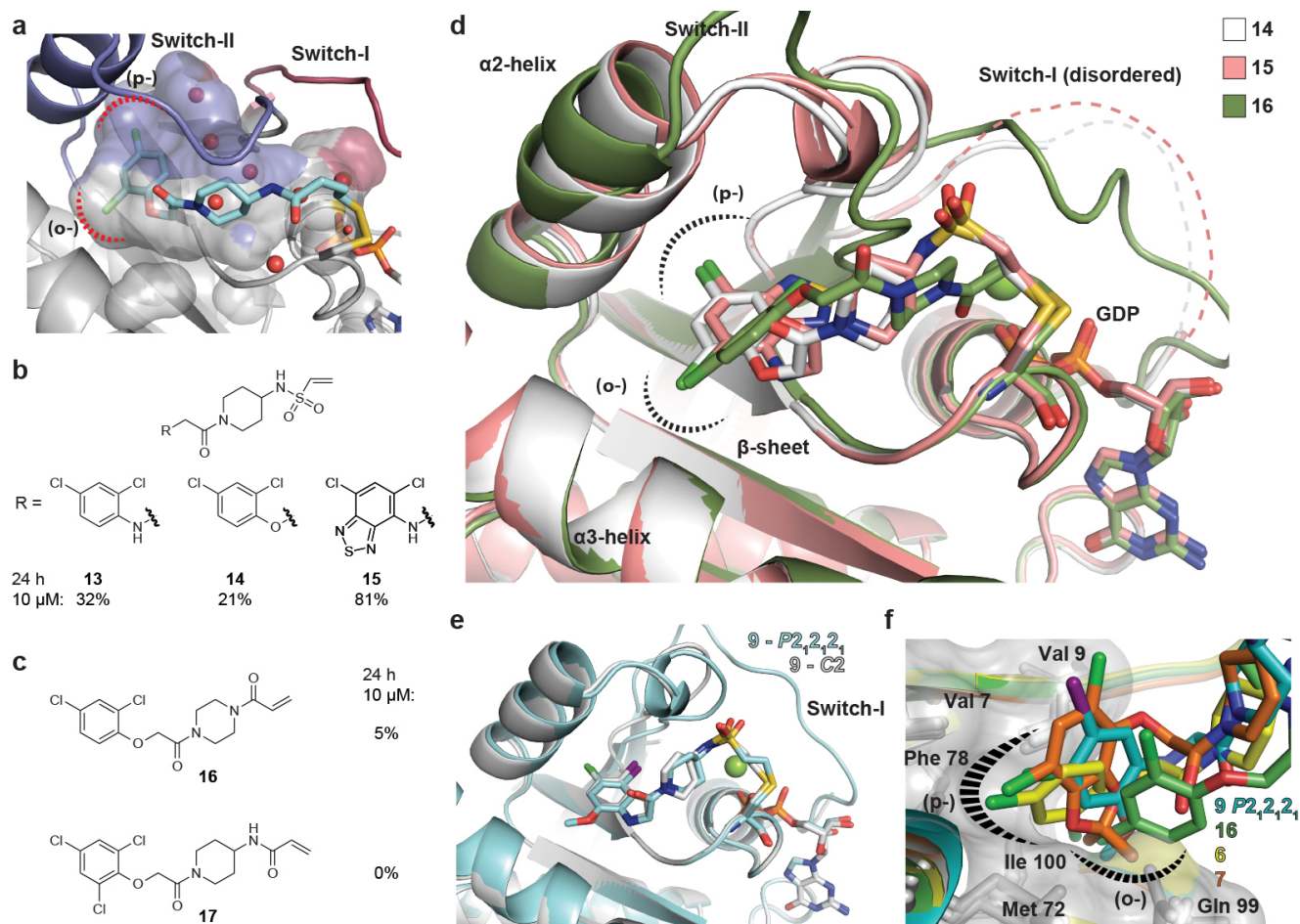
Extended Data Figure 2 | Additional insights into Ras-compound binding and its biochemical effects. a, Compound **6** (cyan) is attached to Cys 12 of K-Ras(G12C) and extends into an allosteric binding pocket beneath switch-II (blue), the S-IIP. The binding pocket in K-Ras (surface representation of the protein shown) fits **6** tightly and includes hydrophobic sub-pockets (dashed lines). An extension of the pocket is occupied by water molecules (red spheres) and might provide space for modified compound analogues. **b–d**, X-ray crystallographic studies of K-Ras(G12C) bound to several additional electrophilic analogues (**14**, **15** and **16**, respectively) reveal a similar overall binding mode. All compounds follow a similar trajectory from Cys 12 into S-IIP but show some variability in the region of the piperidine/piperazine. The respective switch-I regions of the protein can be disordered. **e**, Overlay of the two different crystal forms of K-Ras(G12C) bound to **9** (space group $C2$ (grey) and $P2_12_12_1$ (cyan)) is shown. The ligand orientation and conformation shows minimal changes, whereas switch-II of the protein appears disordered in the $C2$ form and atypical in the $P2_12_12_1$ form. **f**, An overlay for several compounds including the disulphide **6** is shown (**16**-green, **6**-yellow, **7**-orange, **9**-cyan). Key hydrophobic residues are labelled and hydrophobic interaction between the compounds and the (p-) or (o-) sub-pockets are indicated by dashed lines.
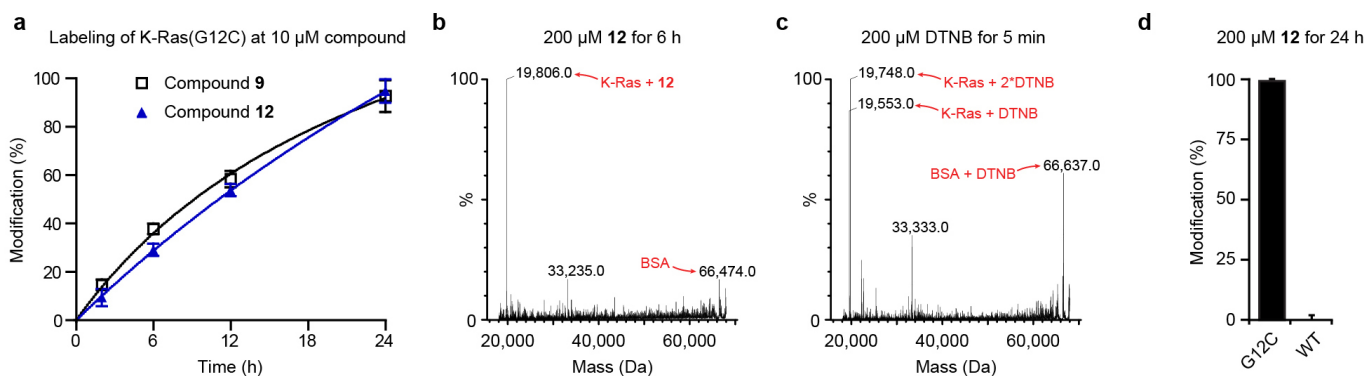
Extended Data Figure 3 | Analysis of compound labelling rate and *in vitro* specificity. a, Percentage modification of K-Ras(G12C) by compounds **9** and **12** over time ($n = 3$, error bars denote s.d.). **b**, Selective single labelling of K-Ras(G12C) by compound **12** in the presence of BSA. **c**, Quantitative single labelling of BSA and multiple labelling of K-Ras(G12C) by DTNB. **d**, Comparison of modification of K-Ras(G12C) and wild-type by **12** ($n = 3$, error bars denote s.d.).
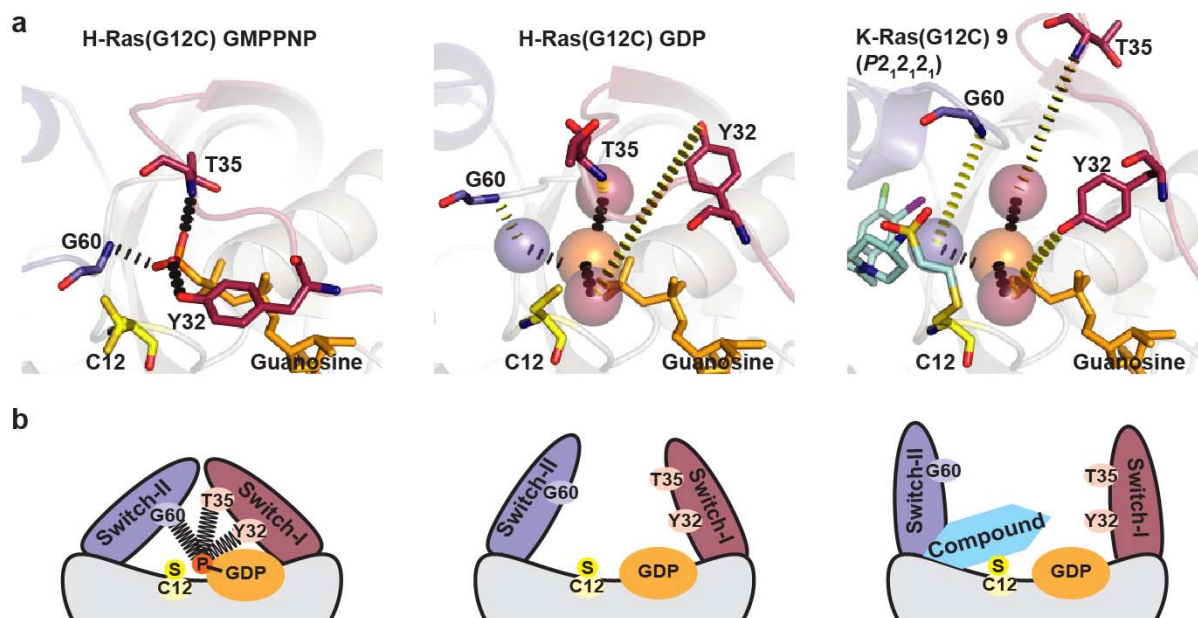
Extended Data Figure 4 | Comparison of active conformation and compound bound form of Ras. a, X-ray crystal structure of the active conformation of H-Ras(G12C) with GMPPNP shows interactions of the γ-phosphate with key residues (Tyr 32, Thr 35 and Gly 60) that hold switch-I (red) and switch-II (blue) in place. The inactive GDP-bound structure of H-Ras(G12C) reveals the absence of these key interactions and increased distances between these residues and the position of the γ-phosphate (positions from GMPPNP structure indicated by spheres) coinciding with large conformational changes in both switch regions. In the $P2_12_12_1$ crystal form of **9** bound to K-Ras(G12C) GDP switch-I is ordered (often disordered by compounds, see Extended Data Table 4), but the structure shows displacement of the γ-phosphate-binding residues beyond their positions in the inactive state. **b**, As indicated by the X-ray structures, removal of the γ-phosphate leads to relaxation of the 'spring-loaded' Ras-GTP back to the GDP state, with opening of switch-II. Compound binding moves switch-II even further away and interferes with GTP binding itself.
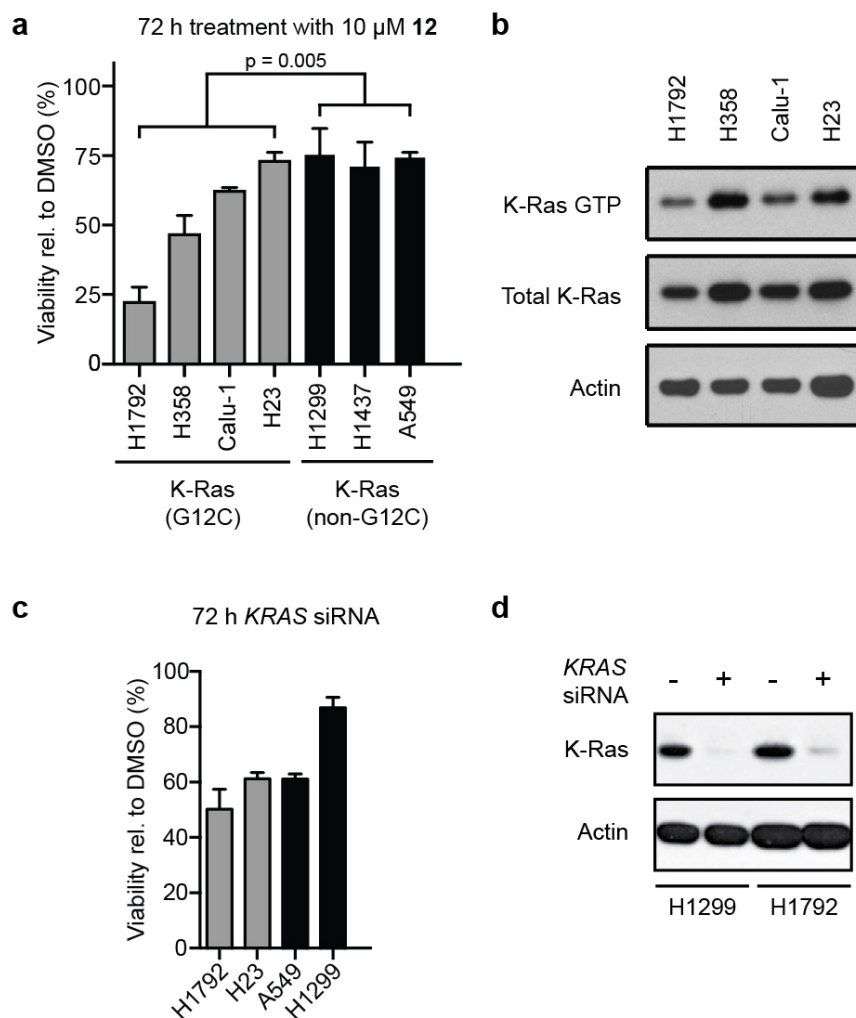
Extended Data Figure 5 | Inhibitor sensitivity, K-Ras GTP levels and K-Ras dependency of lung cancer cell lines. a, Percentage viability after treatment for 72 h with **12** relative to DMSO ($n = 3$ biological replicates, error bars denote s.e.m.). **b**, K-Ras GTP levels determined by incubating lysates with glutathione S-transferase (GST)-tagged RBD (Ras-binding domain of C-Raf) immobilized on glutathione beads ($n = 3$ biological replicates). **c**, Viability of cell lines evaluated 72 h after transfection with *KRAS* siRNA ($n = 3$ biological replicates). **d**, K-Ras immunoblot showing knockdown after *KRAS* siRNA ($n = 3$ biological replicates).
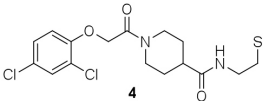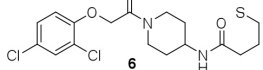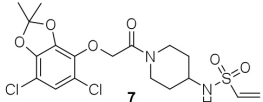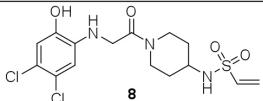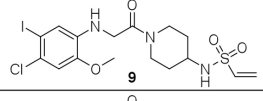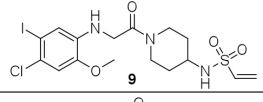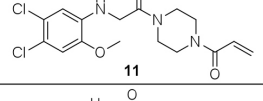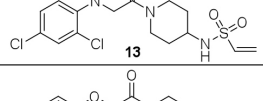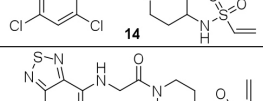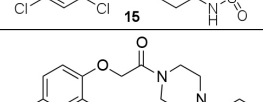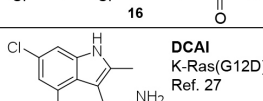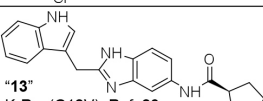
Extended Data Table 1 | Hit fragments and percentage modification from the primary tethering screen

Fragment structure	Fragment number	Percent Modification
	2C10	60%
	2D04	60%
	2D05	60%
	2E07	70%
	3C09	60%
	4C09	60%
	5B03	60%
	5F10	65%
	6H05	95%

Extended Data Table 2 | Overview of obtained and previously published co-crystal structures and their respective compound–protein binding interfaces

Compound	PDB-code	Space group Unit cell	Compound-protein interface ($Å^2$)	Solvent Accessible Surface Area/ Buried Surface Area in structure ($Å^2$)
4	4LV6	$P1$ (33,39,63Å) (77,81,77°)	421	601/531 (chain A)
6	4LUC	$P1$ (33,39,62Å) (78,82,78°)	432	629/548 (chain A)
7	4M1O	$C2$ (68,84,86Å) (90,111,90°)	407	625/489 (chain B)
8	4LYF	$C2$ (68,84,87Å) (90,111,90°)	359	565/430 (chain B)
9	4LYH	$C2$ (68,84,87Å) (90,111,90°)	408	620/507 (chain B)
9	4LYJ	$P2_12_12_1$ (39,43,88Å) (90,90,90°)	412	632/517 (chain A)
11	4M21	$C2$ (68,84,87Å) (90,111,90°)	297	574/372 (chain B)
13	4M1S	$C2$ (68,84,86Å) (90,111,90°)	346	580/442 (chain B)
14	4M1T	$C2$ (68,83,85Å) (90,111,90°)	355	579/444 (chain B)
15	4M1Y	$C2$ (68,84,87Å) (90,111,90^2)	387	599/480 (chain B)
16	4M22	$C2$ (71,83,87Å) (90,109,90°)	324	535/399 (chain B)
DCAI K-Ras(G12D) Ref. 27	4DST	$H3$ (79,79,78.6Å) (90,90,120°)	204	403/252
"13" K-Ras(G12V), Ref. 28	4EPY	$P2_12_12_1$ (39,41,92Å) (90,90,90°)	301	606/370

Extended Data Table 3 | Extent of labelling after 24 h at 10 μM inhibitor

Vinyl sulphonamide inhibitors	Modification (%)		Acrylamide inhibitors	Modification (%)
7	50		10	14
8	87		11	28
9	100		12	100
13	21		16	5
14	32		17	0
15	81			

Extended Data Table 4 | Increased distance (Å) between position-12 Cα and Gly 60 Cα correlates with disordering of switch-I

GDP-bound	12 Cα to 60 Cα distance (Å)	Switch-I	Metal ion
WT	8		Mg
H-Ras(S17N)*	8.1	disordered	Ca
16	8.3		Mg
4	9		Ca
6	9.1		Ca
13	11.1	disordered	-
9 ($P2_12_12_1$)†	11.2	atypical	Mg
14	11.6	disordered	-
8	11.9	disordered	-
15	12	disordered	-
9 ($C2$)†	12.7	disordered	-
7	12.8	disordered	-
11	switch-II disordered	disordered	-

GTP-bound	12 Cα to 60 Cα distance (Å)	Switch-I	Metal ion
H-Ras(G12C)	3.8		Mg
K-Ras(WT)‡	3.9		Mg
Rap1A with CRAF RBD§	3.9		Mg

* PDB accession 3LO5.
† Compound **9** co-crystalized in two different space groups, $P2_12_12_1$ and C_2.
‡ PDB accession 3GFT.
§ PDB accession 1C1Y.

LETTER

Flavin–mediated dual oxidation controls an enzymatic Favorskii-type rearrangement

Robin Teufel[1]*, Akimasa Miyanaga[1]*, Quentin Michaudel[2]*, Frederick Stull[3]*, Gordon Louie[4], Joseph P. Noel[4], Phil S. Baran[2], Bruce Palfey[3,5] & Bradley S. Moore[1,6]

Flavoproteins catalyse a diversity of fundamental redox reactions and are one of the most studied enzyme families[1,2]. As monooxygenases, they are universally thought to control oxygenation by means of a peroxyflavin species that transfers a single atom of molecular oxygen to an organic substrate[1,3,4]. Here we report that the bacterial flavoenzyme EncM[5,6] catalyses the peroxyflavin-independent oxygenation–dehydrogenation dual oxidation of a highly reactive poly(β-carbonyl). The crystal structure of EncM with bound substrate mimics and isotope labelling studies reveal previously unknown flavin redox biochemistry. We show that EncM maintains an unexpected stable flavin-oxygenating species, proposed to be a flavin-N5-oxide, to promote substrate oxidation and trigger a rare Favorskii-type rearrangement that is central to the biosynthesis of the antibiotic enterocin. This work provides new insight into the fine-tuning of the flavin cofactor in offsetting the innate reactivity of a polyketide substrate to direct its efficient electrocyclization.

The antibiotic enterocin (Fig. 1, compound **1**) is produced by various streptomycete bacteria[7] and contains a unique tricyclic caged core. Nearly 40 years ago, isotope labelling studies suggested the involvement of a rare oxidative Favorskii-type rearrangement during its biosynthesis[8].

More recently, the discovery, expression, and biochemical analyses of the *Streptomyces maritimus* enterocin biosynthetic gene cluster including *in vitro* reconstitution of the metabolic pathway showed further involvement of the type II polyketide synthase EncABC and the NADPH-dependent reductase EncD[6,7,9] (Fig. 1). Although type II polyketide synthase pathways typically yield polycyclic aromatic products such as the antibiotic tetracycline and the anticancer agent doxorubicin[10], aromatic polyketides called wailupemycins are formed only as minor products of the enterocin biosynthetic pathway[7]. Remarkably, the FAD-dependent 'favorskiiase' EncM proved to be singly responsible for interruption of the more typical polycyclic aromatization of the poly(β-carbonyl) chain to direct the generation of the rearranged desmethyl-5-deoxyenterocin (**2**)[5,6]. Until now, detailed mechanistic studies of EncM have been hampered by the inherently high reactivity of the proposed EncM substrate, a putative acyl carrier protein (ACP)-bound C7,O4-dihydrooctaketide intermediate (EncC-octaketide; **3**). To overcome this experimental limitation we employed synthetic substrate analogues (for synthesis see Supplementary Information), including the untethered C7,O4-dihydrotetraketide **4**, for structure–function analyses of recombinant EncM.
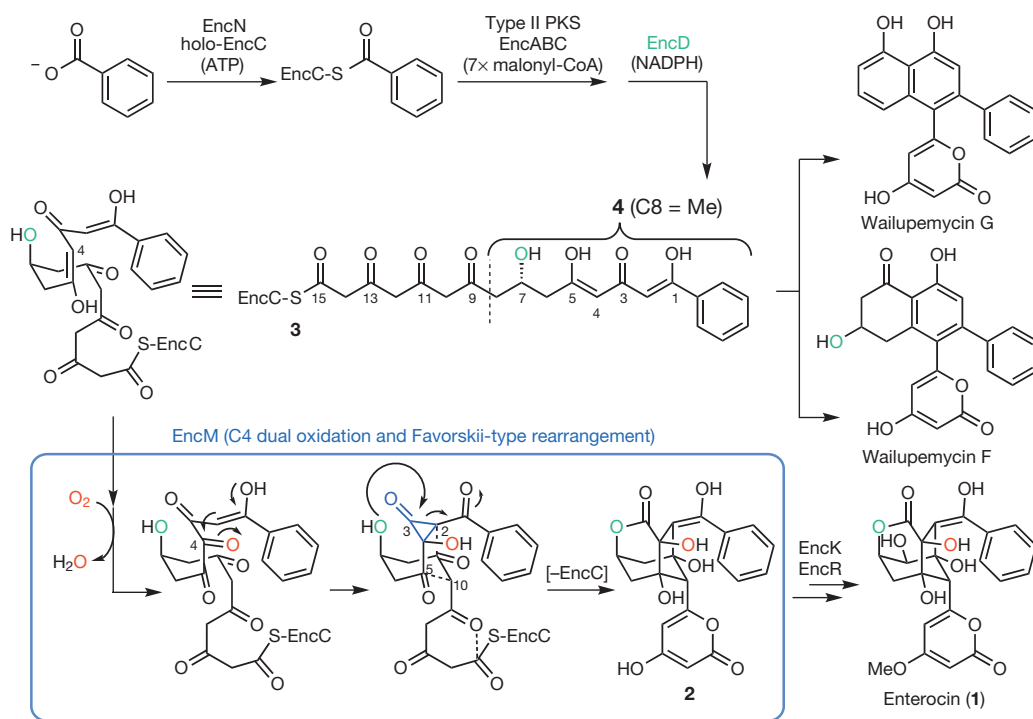

Figure 1 | Overview of the *Streptomyces maritimus* enterocin biosynthetic pathway and proposed EncM catalysis. The ACP EncC is primed with benzoate by ligase EncN, followed by seven iterative type II polyketide synthase (EncAB)-catalysed elongation steps by decarboxylative Claisen condensations with malonyl-CoA. The ketoreductase EncD probably forms the (*R*)-7-hydroxyl group during elongation. The linear (*R*)-C7,O4-dihydrooctaketide (**3**) can cyclize to various wailupemycins (for example G and F), whereas in the presence of EncM it is preferentially converted into desmethyl-5-deoxyenterocin (**2**). Final pathway steps leading to enterocin (**1**) are catalysed by EncR and EncK. EncM catalysis (blue box) involves dual oxidation at C4 (see Fig. 3b) and a Favorskii-type rearrangement, followed by aldol condensations and heterocycle formation (dashed lines). Functional studies of EncM were conducted with the substrate analogue **4**.

[1]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA. [2]Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [3]Program in Chemical Biology, University of Michigan, Ann Arbor, Michigan 48109, USA. [4]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, Jack H. Skirball Center for Chemical Biology and Proteomics, La Jolla, California 92037, USA. [5]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [6]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

Several crystal structures of FAD-bound EncM were determined at resolutions up to 1.8 Å by molecular replacement against 6-hydroxy-D-nicotine oxidase (6HDNO) from *Arthrobacter nicotinivorans*[11] (Fig. 1 and Supplementary Table 1). Structurally, EncM shows greater architectural similarity to flavin dehydrogenases than to oxygenases such as 6HDNO (33% sequence identity for 444 equivalent amino acid residues; 2.2 Å root mean squared deviation (r.m.s.d.) for Cα atoms; Z-score = 46.4), glucooligosaccharide oxidase[12] (31% sequence identity for 415 equivalent residues; 2.3 Å r.m.s.d.; Z-score = 44.1) or aclacino-mycin oxidoreductase[13] (37% sequence identity for 316 equivalent residues; 2.5 Å r.m.s.d.; Z-score = 40.6). In contrast to these monomeric dehydrogenases, EncM exists as homodimer in crystal form and in solution (Fig. 2a and Supplementary Fig. 1). The monomeric subunits of the homodimer show high structural similarity (0.19 Å r.m.s.d. for Cα atoms), and each contains distinct domains for substrate binding (residues 211–418) and FAD binding (residues 2–210 and 419–461). The FAD-binding domain sequesters the ADP-ribosyl of the flavin cofactor, and the reactive isoalloxazine core resides at the interface between the substrate and cofactor domains (Fig. 2a, b). As previously observed in 6HDNO, the flavin is covalently linked to EncM through the C8-methyl group of the isoalloxazine ring system and a histidine residue (His 78) (Fig. 2b).

Structure comparisons with homologous flavin-dependent enzymes emphasized the unusually elongated L-shaped EncM ligand-binding tunnel that extends about 30 Å from the surface to a hydrophobic pocket at its base. This orthogonally arranged two-room tunnel is complementary to the shapes of the ACP-derived phosphopantetheine arm, the octaketide chain and the terminal benzene moiety of **3** (Fig. 2b and Supplementary Fig. 2). The entrance of the tunnel of EncM sits near the dimer interface and adjacent to a surface-exposed basic patch formed

by a few positively charged residues, including Arg 107 and Arg 210, from the dyad-related monomer (Fig. 2a). This positively charged region of EncM is complementary to the decidedly negative surface area of ACPs[14], which is indicative that EncC[7] presents elongated polyketide intermediates to EncM through protein–protein interactions to limit deleterious side reactions of the highly reactive poly(β-carbonyl) chain. Support for the close association of EncM and EncC was obtained by protein–protein computational docking simulation with an EncC homology model (Supplementary Fig. 3). Moreover, disruption of the positive surface area of the EncM dimer with the EncM-R210E mutant resulted in about 40% of the relative activity of native EncM (Supplementary Fig. 4).

To explore the interaction of EncM with the polyketide reactant, we co-crystallized the enzyme with substrate analogues harbouring the benzene moiety of **3** (Supplementary Table 1). The resulting SIGMAA-weighted $F_o - F_c$ electron-density difference maps clearly indicated mimetic binding to the active site, although elevated B-factors and incomplete occupancy (for example roughly 33 Å² and 0.8, respectively, for substrate **4**) caused slightly disordered electron densities (Fig. 2c and Supplementary Fig. 5). Binding occurred with little overall structural perturbation to the EncM polypeptide backbone (for example, 0.14 Å r.m.s.d. for **4**) and no significant backbone or side-chain displacements in the binding region. The terminal benzene group sits at the hydrophobic end of a long tunnel and forms aromatic–aromatic interactions with Tyr 150 and Trp 152 and van der Waals interactions with Leu 357. It is likely that the enol at C1 engages in hydrogen bonding with O4 of the flavin (2.3 Å), whereas the C3 ketone twists away from the flavin and may accept a hydrogen bond from the side chain of Glu 355 (3.2 Å) and possibly from Tyr 249 (3.5 Å). Mutagenesis of these residues confirmed their importance
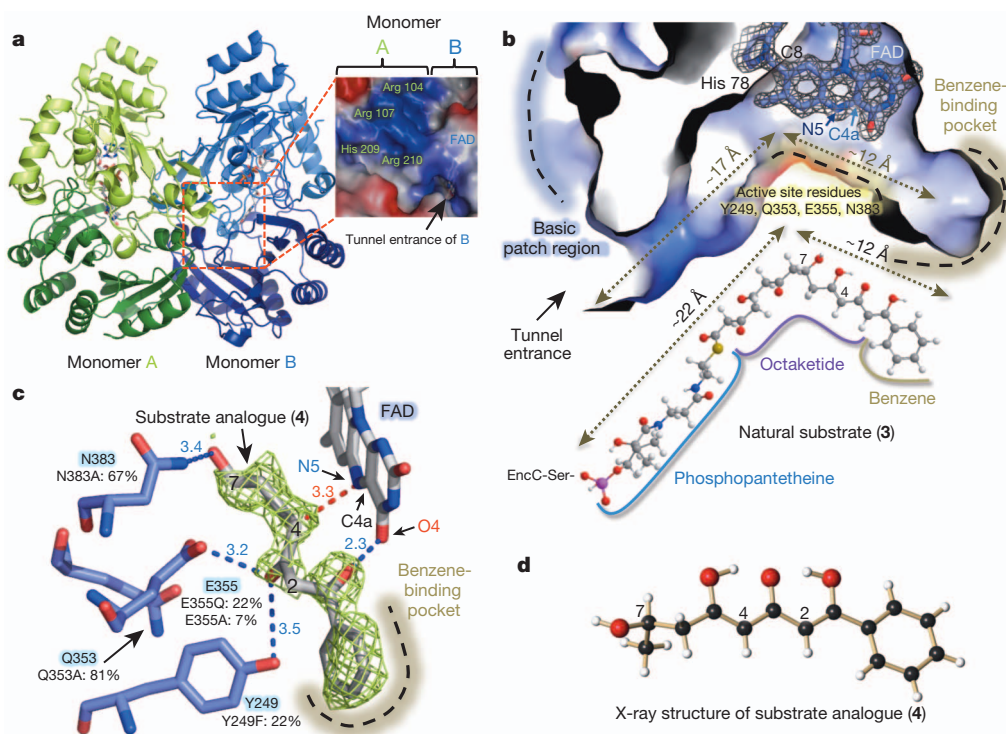

Figure 2 | Crystal structure of EncM. a, Homodimeric EncM shown as a ribbon diagram (with flavin cofactors as colour-coded stick model). Monomeric subunits are coloured in green and blue, with darker shades of each highlighting the substrate-binding domains and lighter shades emphasizing the flavin-binding domains. The basic patch abutting the active-site tunnel entrance (dashed red box) is magnified (blue and red colours indicate positive and negative charges, respectively). **b**, Sliced-away interior view of the EncM substrate tunnel, showing a covalent link between His 78 and FAD (shown is the SIGMAA-weighted $2F_o - F_c$ electron density map contoured at 2.0σ).

The natural substrate **3** is shown below. Approximate lengths of the tunnel and substrate are indicated. **c**, SIGMAA-weighted $F_o - F_c$ difference map of EncM co-crystallized with **4** calculated with the ligand omitted, contoured at 2.0σ around modelled **4**. Hydrogen-bonding interactions are indicated by blue dashed lines. The red dashed line shows the distance (in ångströms) from the site of oxidation to the reactive N5 of FAD. Normalized activities of active site mutants are shown (native EncM = 100%). **d**, X-ray structure of the chemically synthesized substrate analogue **4**.

for EncM activity (Fig. 2c). In particular, the putative C7-hydroxyl group of 4 resides at the elbow of the L-shaped two-room tunnel and ostensibly serves as the pivot point in the natural substrate 3. The mutually orthogonal sections of the EncM ligand-binding pocket separate the C1–C6 triketide head from the C8–C15 pantothenate-linked tetraketide tail to uncouple the reactivity of the entire C1–C16 poly(β-carbonyl) chain. This chemical and structural disconnection prevents kinetically facile but unwanted cyclization–aromatization reactions, and instead favours the EncM-mediated oxidative Favorskii-type rearrangement (Fig. 2b).

We propose that EncM performs a dual oxidation of 3 at C4 to effectively convert a 1,3-diketone to a 1,2,3-triketone. In this mechanistic model, C4 is now set up to undergo a facile electrophilic cyclization with C2 to trigger the proposed Favorskii-like rearrangement (Fig. 1). Typical flavin oxygenases are initially reduced with NAD(P)H to enable the capture of O_2 by reduced flavin (Fl_{red}), generating the flavin-C4a-peroxide oxygenating species[4]. EncM, however, lacks an NAD(P)H-binding domain and functions in the absence of a flavin reductase[6], raising questions surrounding the oxidative mechanism of EncM.

To gain further insight into the EncM chemical mechanism, we analysed the in vitro reaction of EncM with either racemic or enantio-merically pure 4 by reversed-phase high-performance liquid chromatography (HPLC) and ultraviolet–visible spectroscopy. We found that 4 was converted in the absence of NAD(P)H into diastereomeric products 5 and 5′ without detectable intermediates (Fig. 3a). Through comprehensive NMR and mass spectrometric analyses together with chemical synthesis (see Supplementary Information), we identified 5 and 5′ as ring-opened derivatives of the expected enterocin-like lactone 6 (Fig. 3b). Circular dichroism experiments proved that the configuration of 4 is maintained during the transformation (see Supplementary Information). We reasoned that a facile hydrolytic retro-Claisen ring cleavage[15,16] of 6 occurs after an oxidative Favorskii-type rearrangement and lactonization (Fig. 3b, step VII) that is probably responsible for the racemization of C4. This proposed reaction was further substantiated by the observation that glycerol also effectuates the ring opening to form 7 and 7′ (Fig. 3a and Supplementary Figs 6 and 7). During enterocin biosynthesis this reaction is probably prevented through aldol condensations with the remainder of the ketide chain (Fig. 1).
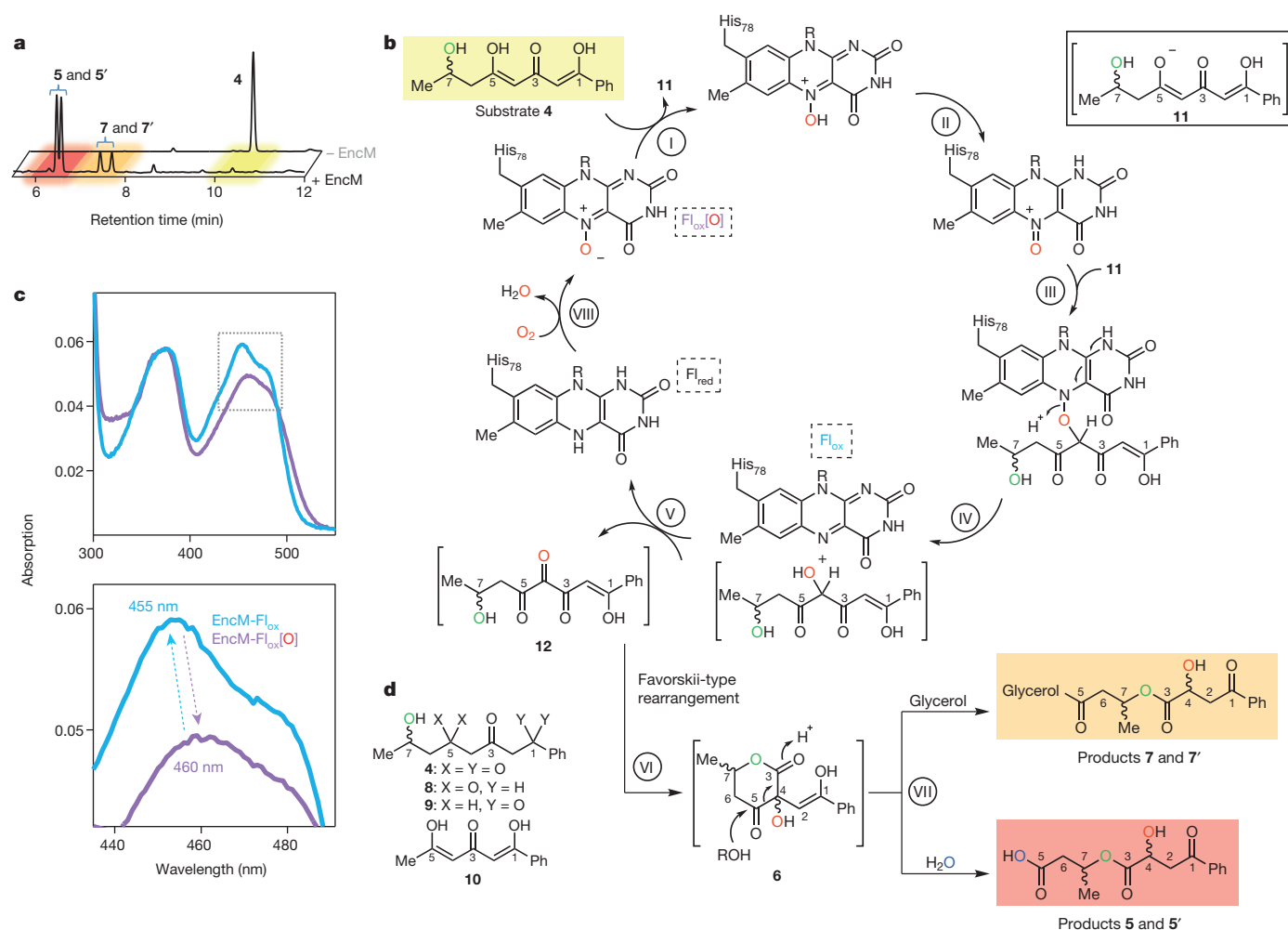

Figure 3 | Proposed EncM mechanism and spectral features of the flavin cofactor catalytic states. a, Reversed-phase HPLC analysis (absorption detection at 254 nm) of enzymatic assays showing substrate analogue 4 (upper lane; control assay without EncM) and diastereomeric product pairs 5/5′ and 7/7′ (lower lane; after incubation with EncM). The colour code refers to b. Products 7/7′ were observed only in the presence of glycerol (here 20% v/v). No intermediates could be detected. b, Proposed catalytic mechanism of EncM involving substrate oxygenation by means of a flavin-N5-oxoammonium species. The resultant electrophilic C4-ketone of 12 triggers the Favorskii-type rearrangement and lactone formation (see Fig. 1 for the detailed analogous reactions during the natural biosynthesis of enterocin), while the formed Fl_{red}

reacts with O_2 and restores the N5-oxide. The stepwise dual oxidation is supported by anaerobic single-turnover experiments (Supplementary Fig. 16). The C7-hydroxyl group is shown in green, and oxygen atoms derived from O_2 and H_2O are colour-coded red and blue, respectively. Roman numerals indicate reaction steps as discussed in the main text. c, Ultraviolet–visible spectra of the oxidized flavin of EncM as isolated ($Fl_{ox}[O]$, catalytically active, purple curve) and after multiple substrate turnovers (Fl_{ox}, catalytically inactive, blue curve). Molar absorption coefficients were $\varepsilon_{450} = 11{,}900\ M^{-1}\ cm^{-1}$ for EncM-Fl_{ox} and $\varepsilon_{460} = 9{,}600\ M^{-1}\ cm^{-1}$ for EncM-$Fl_{ox}[O]$. d, Compounds used for structure–activity relationship analyses.

Notably, the C1 and C5 deoxo-substrate analogues **8** and **9**, respectively, were not transformed by EncM, whereas the dehydroxy-substrate **10** (see Fig. 3d or Supplementary Fig. 5 for compound structures) was converted into multiple unstable products that were not further characterized. This series of structure–activity relationships revealed that the triketone motif (C1–C6) is essential for catalysis and suggested that the C7-hydroxyl group is critical for spatial and temporal control of the EncM catalysed reaction.

The monooxygenase activity of EncM was evaluated by following the incorporation of oxygen atoms from $^{18}O_2$ into **5/5′** and **7/7′** at C4. In contrast, isotope labelling from $H_2^{18}O$ was only associated with the non-enzymatic retro-Claisen cleavage of **6** to **5/5′** (Supplementary Figs 8 and 9). These measurements suggest that lactone formation during enterocin biosynthesis is controlled by the C7-hydroxyl group by means of direct intramolecular attack (Fig. 1). Further support for this biosynthetic model came from the structure analysis of the EncM ligand-binding tunnel that can only accommodate the (*R*)-enantiomer of **3** (Supplementary Fig. 10), which is consistent with the observed retention of the C4-hydroxyl configuration in the final product enterocin (Fig. 1).

EncM became inactivated after several turnovers (Supplementary Fig. 11). Moreover, the oxidized flavin cofactor of inactivated EncM (EncM-Fl_{ox}) showed distinct, stable changes in the ultraviolet–visible spectrum (Fig. 3c). We speculated that these spectral perturbations are caused by the loss of an oxygenating species maintained in the enzyme's active state. This species, 'EncM-Fl_{ox}[O]', is largely restored at the end of each catalytic cycle (Fig. 3b), thereby providing an explanation for the innate monooxygenase activity of EncM in the absence of exogenous reductants. We excluded the participation of active-site residues in harbouring this oxidant by using site-directed mutagenesis and by showing that denatured EncM retained the Fl_{ox}[O] spectrum (Supplementary Fig. 12). We therefore focused on the flavin cofactor as the carrier of the oxidizing species. On the basis of the spectral features of EncM-Fl_{ox}[O], we ruled out a conventional C4a-peroxide[17,18]. Moreover, Fl_{ox}[O] is extraordinarily stable (no detectable decay for more than 7 days at 4 °C) and thus is vastly longer lived than even the most stable flavin-C4a-peroxides described so far ($t_{1/2} \leq$ 30 min at 4 °C (refs 19, 20)).

To further test the possible intermediacy and catalytic role of EncM-Fl_{ox}[O], we reduced the flavin cofactor anaerobically and showed that only flavin reoxidation with molecular oxygen restored the EncM-Fl_{ox}[O] species. In contrast, anoxic chemical reoxidation generated catalytically inactive EncM-Fl_{ox} (Supplementary Fig. 13a). Notably, EncM reoxidized with $^{18}O_2$ formed EncM-Fl_{ox}[^{18}O], which converted **4** to [^{18}O]**5/5′** with 1:1 stoichiometry of Fl_{ox}[^{18}O] to [^{18}O]**5/5′** (Supplementary Fig. 13b). The collective structure–function analyses reported here currently support the catalytic use of a unique flavin-oxygenating species that is consistent with a flavin-N5-oxide. This chemical species was introduced more than 30 years ago as a possible intermediate in flavin monooxygenases[21,22] before the conventional C4a-peroxide model was accepted experimentally. Crucially, spectrophotometric comparison of chemically synthesized flavin-N5-oxide and EncM-Fl_{ox}[O] revealed many of the same spectral features[23], and both can be chemically converted to oxidized flavin (Supplementary Fig. 12). Moreover, consistent with an N-oxide, EncM-Fl_{ox}[O] required four electrons per flavin cofactor to complete reduction in dithionite titrations, whereas EncM-Fl_{ox} required only two (Supplementary Fig. 14). We could not observe this flavin modification crystallographically (see Fig. 2b), presumably as a result of X-radiation-induced reduction[24] of the flavin-N5-oxide, which is highly prone to reduction[23].

We propose that during EncM catalysis, the N5-oxide is first protonated by the hydroxyl proton of the C5-enol of substrate **4** (Fig. 3b, step I). Despite the generally low basicity of N-oxides, the proton transfer is probably enabled by the high acidity of the C5 enol and its appropriate positioning 3.3 Å from the N5 atom of the flavin (Fig. 2c). After protonation, tautomerization of the N5-hydroxylamine would lead to the electrophilic oxoammonium (step II). Subsequent oxygenation of substrate enolate **11** by the oxoammonium species may then occur by one of several possible routes (Supplementary Fig. 15), yielding Fl_{ox} and a C4-hydroxylated intermediate (steps III and IV). Fl_{ox}-mediated dehydrogenation of the introduced alcohol group then produces the C4-ketone **12** and Fl_{red} (step V). Anaerobic single turnover experiments with **4** support this reaction sequence (Supplementary Fig. 16). Finally, **12** would undergo the Favorskii-type rearrangement (step VI) and retro-Claisen transformation (step VII) to yield the observed products **5/5′** or **7/7′**, while the reduced cofactor Fl_{red} reacts with O_2 to regenerate EncM-Fl_{ox}[O] and thus prime the enzyme for the next catalytic cycle (steps VIII). However, alternative mechanisms are also plausible (Supplementary Fig. 17). This extraordinary flavin cofactor-mediated dual oxidation vaguely resembles the role of flavins in the scarce 'internal monooxygenases' (EC 1.13.12) that also use their substrate as an electron donor[25].

Here we provide the first in-depth investigation of an enzymatic oxidation-induced Favorskii-type rearrangement. The exceptionally reactive poly(β-carbonyl) substrate requires EncM to direct the reaction along a defined mechanistic trajectory by sequestration of reactants from bulk solvent, spatial separation of reactive functional groups, rapid 'one-step' generation of a new electrophilic centre, and expulsion of solvent from the active site to prevent retro-Claisen ring cleavage. The discovery that EncM uses a stable flavin-N5-oxide for oxygenation rather than the universally accepted flavin peroxide suggests that this species may have been overlooked in the flavin biochemical literature. Further studies are under way to explore the factors that govern enzymatic formation of the flavin-N5-oxide. In short, the archetypal dual oxidase EncM employs unexpected oxidative flavin biochemistry for the NAD(P)H-independent processing of extremely reactive polyketides.

METHODS SUMMARY

Amino-terminal octahistidine-tagged EncM from *S. maritimus* was produced heterologously in *Escherichia coli* BL21 (DE3) and purified by means of Ni^{2+}-affinity chromatography. For crystallization, the EncM His-tag was removed and the protein was further purified by ResourceQ anion-exchange chromatography. Substrate analogues and flavin-N5-oxide were acquired through chemical synthesis. Site-directed mutagenesis was conducted with the QuikChange site-directed mutagenesis kit (Stratagene), using self-constructed primers.

The activities of wild-type EncM and EncM-R210E were assayed using the fully reconstituted enzyme set as reported previously[6]. Other EncM assays were conducted at 22 °C and pH 7.5, using HEPES-Na^+ buffer, 150–300 mM NaCl and at least 10% (v/v) glycerol. Products were separated and purified by reverse-phase HPLC with optical detection at 254 nm using a Sync Polar RP column with an ammonium acetate-buffered (pH 5.0) acetonitrile gradient. A 6230 Accurate-Mass TOF–MS system (Agilent) was used for mass spectrometric measurements. NMR spectra were recorded on Bruker DRX-600 and AMX-400 instruments. Ultraviolet–visible spectra were obtained with a Cary 50 UV-Vis spectrophotometer (Agilent). A Perkins-Elmer 341 polarimeter and an Aviv circular dichroism spectrometer were used for optical rotation and circular dichroism spectroscopy measurements, respectively.

Crystals of EncM were grown from a 1:1 mixture of protein solution (5 mg ml^{-1} in 10 mM TES-Na^+ pH 7.7 containing 10% (v/v) glycerol) and a reservoir solution (2 mM dithiothreitol, 0.1 M HEPES-Na^+ pH 7.5, 0.2 M calcium acetate, 20% PEG3350) using hanging-drop vapour diffusion method at 4 °C. For co-crystallization the enzyme was incubated with the substrate mimic (2 mM) before being mixed with the reservoir solution. The crystals were stored in 25% (v/v) glycerol until X-ray data collection at the Advanced Light Source (Berkeley, CA, USA). The initial phases were determined by molecular replacement with 6HDNO (PDB 2BVG) as a search model.

Full Methods and any associated references are available in the online version of the paper.

1. Walsh, C. T. & Wencewicz, T. A. Flavoenzymes: versatile catalysts in biosynthetic pathways. *Nat. Prod. Rep.* **30,** 175–200 (2012).

2. Chaiyen, P., Fraaije, M. W. & Mattevi, A. The enigmatic reaction of flavins with oxygen. *Trends Biochem. Sci.* **37,** 373–380 (2012).
3. Massey, V. Activation of molecular oxygen by flavins and flavoproteins. *J. Biol. Chem.* **269,** 22459–22462 (1994).
4. Palfey, B. A. & McDonald, C. A. Control of catalysis in flavin-dependent monooxygenases. *Arch. Biochem. Biophys.* **493,** 26–36 (2010).
5. Xiang, L., Kalaitzis, J. A. & Moore, B. S. EncM, a versatile enterocin biosynthetic enzyme involved in Favorskii oxidative rearrangement, aldol condensation, and heterocycle-forming reactions. *Proc. Natl Acad. Sci. USA* **101,** 15609–15614 (2004).
6. Cheng, Q., Xiang, L., Izumikawa, M., Meluzzi, D. & Moore, B. S. Enzymatic total synthesis of enterocin polyketides. *Nature Chem. Biol.* **3,** 557–558 (2007).
7. Piel, J. *et al.* Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': evidence for the derailment of an aromatic polyketide synthase. *Chem. Biol.* **7,** 943–955 (2000).
8. Seto, H., Sato, T., Urano, S., Uzawa, J. & Yonehara, H. Utilization of ¹³C–¹³C coupling in structural and biosynthetic studies. VII. The structure and biosynthesis of vulgamycin. *Tetrahedr. Lett.* **48,** 4367–4370 (1976).
9. Hertweck, C. *et al.* Context-dependent behavior of the enterocin iterative polyketide synthase; a new model for ketoreduction. *Chem. Biol.* **11,** 461–468 (2004).
10. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24,** 162–190 (2007).
11. Koetter, J. W. & Schulz, G. E. Crystal structure of 6-hydroxy-D-nicotine oxidase from *Arthrobacter nicotinovorans. J. Mol. Biol.* **352,** 418–428 (2005).
12. Huang, C. H. *et al.* Crystal structure of glucooligosaccharide oxidase from *Acremonium strictum*: a novel flavinylation of 6-S-cysteinyl, 8α-N1-histidyl FAD. *J. Biol. Chem.* **280,** 38831–38838 (2005).
13. Alexeev, I., Sultana, A., Mantsala, P., Niemi, J. & Schneider, G. Aclacinomycin oxidoreductase (AknOx) from the biosynthetic pathway of the antibiotic aclacinomycin is an unusual flavoenzyme with a dual active site. *Proc. Natl Acad. Sci. USA* **104,** 6170–6175 (2007).
14. Crosby, J. & Crump, M. P. The structural role of the carrier protein–active controller or passive carrier. *Nat. Prod. Rep.* **29,** 1111–1137 (2012).
15. Baumann, K., Bacher, M., Damont, A. & Steck, A. Selective transformation of ascomycin into 11-epi-ascomycin. *Tetrahedr. Lett.* **45,** 549–551 (2004).
16. Takikawa, H., Takada, A., Hikita, K. & Suzuki, K. Formation of α-hydroxy-β-diketones through hydroxylation of isoxazolium salts: stereoselective approach to angular cis-diols in polycyclic systems. *Angew. Chem. Int. Edn Engl.* **47,** 7446–7449 (2008).
17. Entsch, B., Ballou, D. P. & Massey, V. Flavin–oxygen derivatives involved in hydroxylation by p-hydroxybenzoate hydroxylase. *J. Biol. Chem.* **251,** 2550–2563 (1976).
18. Entsch, B. & Ballou, D. P. Purification, properties, and oxygen reactivity of p-hydroxybenzoate hydroxylase from *Pseudomonas aeruginosa. Biochim. Biophys. Acta* **999,** 313–322 (1989).
19. Thotsaporn, K., Chenprakhon, P., Sucharitakul, J., Mattevi, A. & Chaiyen, P. Stabilization of C4a-hydroperoxyflavin in a two-component flavin-dependent monooxygenase is achieved through interactions at flavin N5 and C4a atoms. *J. Biol. Chem.* **286,** 28170–28180 (2011).
20. Valton, J., Mathevon, C., Fontecave, M., Niviere, V. & Ballou, D. P. Mechanism and regulation of the two-component FMN-dependent monooxygenase ActVA-ActVB from *Streptomyces coelicolor. J. Biol. Chem.* **283,** 10287–10296 (2008).
21. Rastetter, W. H., Gadek, T. R., Tane, J. P. & Frost, J. W. Oxidations and oxygen transfers effected by a flavin N(5)-oxide. A model for flavin-dependent monooxygenases. *J. Am. Chem. Soc.* **101,** 2228–2231 (1979).
22. Orf, H. W. & Dolphin, D. Oxaziridines as possible intermediates in flavin monooxygenases. *Proc. Natl Acad. Sci. USA* **71,** 2646–2650 (1974).
23. Walsh, C. *et al.* Chemical and enzymatic properties of riboflavin analogues. *Biochemistry* **17,** 1942–1951 (1978).
24. Garman, E. F. & Owen, R. L. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 32–47 (2006).
25. van Berkel, W. J., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* **124,** 670–689 (2006).

Author Contributions R.T., A.M., Q.M., F.S., G.L. and J.P.N. performed research. All authors designed research and analysed data. R.T. and B.M. wrote the paper. R.T., A.M., Q.M. and F.S. contributed equally to this work.

Author Information The EncM sequence is deposited in GenBank under accession number AAF81732.1. The structures for proteins described in this paper have been deposited in the Protein Data Bank under accession numbers 3W8W (apo-EncM), 3W8X (EncM with bound **26**) and 3W8Z (EncM with bound **4**). Data for the crystallized substrate analogues are deposited in the Cambridge Crystallographic Data Centre under accession numbers CCDC 922822 (**4**), CCDC 922821 (**10**) and CCDC 949270 (**26**). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.S.M. (bsmoore@ucsd.edu).

METHODS

Gene cloning, heterologous protein expression, and purification procedures. *Escherichia coli* strain BL21 (DE3) (New England Biolabs) and *Streptomyces lividans* TK24 were used for heterologous protein expression. The enterocin enzymes holo-EncC[26], EncA-EncB[26], EncD[6] and EncN[27] from *Streptomyces maritimus*, and FabD[28] from *Streptomyces glaucescens*, were prepared as His-tagged recombinant proteins as described previously[6,26–28]. The plasmid encoding FabD was provided by K. A. Reynolds. The EncM gene was amplified from pXY200-EncM[6] with the primers 5′-AAAAC**CCATGG**GCAGTTCCCACAGCTCGAC-3′ and 5′-TTTT**GAATTCT**CAGGGGCTGCTCGGG-3′ (NcoI and EcoRI restriction sites are underlined) and then inserted between the NcoI and EcoRI sites of the expression vector pHIS8 (ref. 29). *E. coli* BL21 (DE3) harbouring pHIS8-EncM plasmid was grown at 28 °C in 4 l of lysogeny broth containing 50 μg ml^{-1} kanamycin until D_{600} reached about 0.5. Isopropyl-β-D-thiogalactoside (25 μM) was then added to induce recombinant protein expression under the control of T7 RNA polymerase induced using a modified *lac* promoter. Cells were grown for a further 24 h at 28 °C and harvested by centrifugation. Cell pellets were resuspended in lysis buffer (50 mM sodium phosphate pH 7.7, 300 mM sodium chloride, 10% (v/v) glycerol) supplemented with 10 mM imidazole, and lysed by sonication. After centrifugation, the supernatant was passed over a Ni^{2+}-nitrilotriacetate column connected to a FPLC system. Unbound protein was removed by washing and the N-terminal octahistidine-tagged EncM was then eluted with lysis buffer supplemented with 500 mM imidazole. The protein was desalted and concentrated using PD-10 and Vivaspin 6 (30 kDa exclusion size) columns (both from GE Healthcare), respectively. For crystallization, EncM was further treated with thrombin to remove the His$_8$ tag and subjected to another round of His-trap purification followed by ResourceQ (GE Healthcare) anion-exchange chromatography with a linear gradient from 0 to 1 M NaCl over 30 min in 10 mM TES-Na$^+$ buffer pH 7.7 containing 10% (v/v) glycerol.

Hydrodynamic analysis of EncM by size-exclusion chromatography. EncM protein (0.5 mg) was loaded onto a HiLoad 26/60 Superdex 200 column equilibrated with buffer containing 20 mM TES-Na$^+$ pH 7.5, 0.15 M NaCl and 10% (v/v) glycerol. Eluted protein was observed by monitoring the absorbance at 280 nm. The column was calibrated with Bio-Rad standard proteins (thyroglobulin, 670 kDa; γ-globulin, 158 kDa; ovalbumin, 44 kDa; myoglobin, 17 kDa).

Molar absorption coefficients of EncM-Fl$_{ox}$[O] and EncM-Fl$_{ox}$. A solution of anaerobic dithionite in a gastight syringe was calibrated by titrating a known concentration of flavin mononucleotide to full reduction. The dithionite syringe was transferred to an anaerobic cuvette containing EncM-Fl$_{ox}$ and then titrated with the calibrated dithionite to complete reduction. The amount of dithionite needed to reduce EncM-Fl$_{ox}$ fully was used to determine the molar absorption coefficient (ε) of 11,900 M^{-1} cm^{-1} at 450 nm on the basis of the original absorbance spectrum. Subsequent exposure to O$_2$ led to oxidation of the reduced EncM to EncM-Fl$_{ox}$[O], from which $\varepsilon = 9,600$ M^{-1} cm^{-1} at 460 nm was calculated.

Site-directed mutagenesis. The expression plasmid pHIS8-EncM was used for site-directed mutagenesis with the QuikChange site-directed mutagenesis kit in accordance with the manufacturer's protocol (Stratagene). The following oligonucleotides (and respective complementary primers) were used to obtain the EncM mutants R210E, Y249F, Q353A, E355A, E355Q and N383A, respectively: 5′-GAG TTCGACCTCCACGAGGTCGGGCCCGTC-3′, 5′-CTGACCTGGGCGTTGTT TCTGCGCCTGGCAC-3′, 5′-GCCTCCCCCTTCACTGCGCTCGAACTGCTC TACC-3′, 5′-CCCTTCACTCAGCTCGCACTGCTCTACCTGGG-3′, 5′-CCCT TCACTCAGCTCCAACTGCTCTACCTGGG-3′ and 5′-CGCCGTTCGTGACC GCCCTGGCCGCCGC-3′. The mutations were confirmed by sequence analysis.

Crystallization, structure determination, and refinement. Crystals of EncM were grown from a 1:1 mixture of protein solution (5 mg ml^{-1} in 10 mM TES-Na$^+$ pH 7.7 containing 10% (v/v) glycerol) and a reservoir solution (2 mM dithiothreitol, 0.1 M HEPES-Na$^+$ pH 7.5, 0.2 M calcium acetate, 20% (w/v) PEG3350) using hanging-drop vapour diffusion at 4 °C. For co-crystallization, EncM was incubated with the respective substrate analogues (2 mM) before being mixed with the reservoir solution. The crystals were transferred to the reservoir solution, containing 25% (v/v) glycerol as a cryoprotectant, and flash-frozen in liquid nitrogen until X-ray data collection on beamlines 8.2.1 and 8.2.2 at the Advanced Light Source (Berkeley, CA, USA). All diffraction data were indexed, integrated and scaled with the HKL2000 (ref. 30) or iMosflm[31]. The initial phases were determined by molecular replacement using Molrep[32]. The crystal structure of 6HDNO (PDB 2BVG) was used as a search model, and ARP/wARP[33], Coot[34] and Refmac[35] were used for automatic model building, for visual inspection and manual rebuilding of the model, and for several rounds of energy minimization and individual *B*-factor refinement, respectively. Ramachandran statistics: EncM apo, favoured region 98.0%, allowed region 1.5%, outlier region 0.4%; EncM with **26**, favoured region 98.8%, allowed region 1.1%, outlier region 0.1%; EncM with **4**, favoured region 98.8%, allowed region 1.0%, outlier region 0.2%. The figures were prepared using

Pymol[36]. Occupancies and *B*-factors for EncM-bound substrate analogues were determined with Phenix[37].

Enzyme assays (Fig. 3a and Supplementary Fig. 11). The kinetics for product formation were determined at 22 °C using two replicate assays containing 20 mM HEPES-Na$^+$ pH 7.5, 300 mM NaCl, at least 10% (v/v) glycerol, 0.7 mM **4** and 10 μM EncM. EncM concentrations were adjusted on the basis of the molar absorption coefficient of EncM-Fl$_{ox}$[O] (9,600 M^{-1} cm^{-1}) at 460 nm. Samples were withdrawn sequentially and quenched after 1, 3, 6, 12, 20, 30 and 40 min. To determine native and mutant EncM activities, a final concentration of 3.4 μM of each EncM mutant was incubated with 0.6 mM **4** in 50 mM HEPES-Na$^+$ pH 7.5, 200 mM NaCl, 1 mM NADPH, 10% (v/v) glycerol using three replicate assays. The reactions were quenched after 10 min (when less than 50% of the substrate had been converted) and the products were quantified. All samples described in this section were analysed by HPLC (see below).

EncM flavin oxidation with molecular oxygen ($^{18}O_2$ or $^{16}O_2$) and 2,6-dichlorophenolindophenol (Supplementary Fig. 13). EncM-Fl$_{ox}$[O] (20 μM) active sites were completely reduced in an anaerobic cuvette with sodium dithionite before reoxidation by injection of about 97% $^{18}O_2$ gas (Sigma-Aldrich), about 50% $^{18}O_2$ gas (1:1 mixture of $^{18}O_2$ and $^{16}O_2$) or air. Unreacted O$_2$ was then thoroughly removed by repeated cycles of treatment with vacuum and argon; 100 μM **4** was then added at room temperature. After complete consumption of **4**, protein was removed through filtration and the samples were acidified with 1 M HCl before liquid chromatography–mass spectrometric analysis. Alternatively, EncM was reoxidized anaerobically with the chemical oxidant 2,6-dichlorophenolindophenol instead of O$_2$, producing catalytically inactive EncM-Fl$_{ox}$ (no products were detected after incubation with **4**).

Model docking (Supplementary Fig. 3c). The homology model of EncC was generated by Swiss Model[38] on the basis of the solved structure of the ACP of actinorhodin biosynthesis from *Streptomyces coelicolor* (PDB 1AF8). Docking simulation was performed with the GRAMM-X Protein–protein Docking Web Server[39], using the EncM structure and the EncC homology model. The resulting structure was then energy-minimized with Swiss-model viewer[40].

***In vitro* reconstitution assay with the enterocin PKS (Supplementary Fig. 4).** The activities of EncM and EncM-R210E were assayed using the fully reconstituted *enc* PKS enzyme set as reported previously[6]. The standard mixture contained 1 μM EncA-EncB, 8 μM EncC, 1.5 μM EncD, 2 μM EncM, 0.15 μM EncN, 0.015 μM FabD, 5 mM ATP, 5 mM MgCl$_2$, 5 mM NADPH, 1 mM malonyl-CoA and 0.25 mM benzoic acid in a volume of 100 μl. After incubation at 30 °C for 2 h, the reactions were quenched by the addition of 10 μl of 2 M HCl. The products were then extracted twice with 200 μl of ethyl acetate. The organic extracts were combined and evaporated to dryness. The residual material was resuspended in 30 ml of acetonitrile and analysed by HPLC and LC-ESI mass spectrometry. A Phenomenex C$_{18}$ column (250 mm × 4.6 mm) was used at a flow rate of 1.0 ml min^{-1} with a linear gradient of 5–80% (v/v) acetonitrile in water containing 0.1% (v/v) trifluoroacetic acid over a period of 40 min.

Ultraviolet–visible spectrophotometry (Fig. 3c and Supplementary Figs 12–14). The flavin absorption spectra of purified EncM were analysed with an Agilent Cary 50 UV-Vis spectrophotometer or a Shimadzu UV-2501 PC. Untreated EncM (as isolated from *E. coli*) showed the EncM-Fl$_{ox}$[O] spectrum. After incubation with substrate (and subsequent product removal using a PD-10 column), the spectrum of EncM-Fl$_{ox}$ was observed.

Analytic (Fig. 3a), semipreparative and chiral HPLC. Samples from enzymatic assays were quenched in acidic methanol and centrifuged. The supernatants were analysed by reverse-phase HPLC (1200 series; Agilent) using a Sync Polar RP column 4μ (150 mm × 4.6 mm; ES Industries) with 10% (v/v) acetonitrile as liquid phase buffered in 90% (v/v) 20 mM ammonium acetate pH 5.0. The buffer was gradually exchanged for acetonitrile using a linear gradient from 10% to 95% (v/v) acetonitrile over 15 min at a flow rate of 1 ml min^{-1}. Products were quantified on the basis of D_{254} using a standard curve. Semipreparative reverse-phase HPLC was performed using a Waters 600 controller coupled to a Waters 990 photodiode array detector. Chiral HPLC was performed using a SPD-10A VP Shimadzu system.

Mass spectrometry. Samples were purified by HPLC as described above and then analysed with high-resolution electrospray ionization MS (positive mode) using a 6230 Accurate-Mass TOF MS system (Agilent). Alternatively, a 1290 Infinity LC system coupled to a 6530 Accurate-Mass Q-TOF MS system (both from Agilent) was employed. HPLC was conducted using a Phenomenex Luna 5μ C18E (2) column (150 mm × 4.6 mm) with an acetonitrile gradient of 10–90% (v/v) over 25 min in 0.1% (v/v) formic acid. For synthesized **5** and **5′** and intermediates, high-resolution mass spectra were recorded on an Agilent LC/MSD TOF mass spectrometer by electrospray ionization time-of-flight (ESI-TOF) reflectron experiments.

NMR spectroscopy. NMR spectra were recorded on Bruker DRX-600 and AMX-400 instruments and were calibrated using residual undeuterated solvent as an

internal reference (CHCl$_3$ at 7.26 p.p.m. ^{1}H-NMR, 77.16 p.p.m. ^{13}C-NMR). The following abbreviations were used to explain NMR peak multiplicities: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; br, broad.

Optical rotations and circular dichroism spectroscopy. Optical rotations were obtained on a Perkin-Elmer 341 polarimeter. Circular dichroism spectroscopy measurements were obtained on an Aviv circular dichroism spectrometer model 62DS.

Chemical syntheses. See Supplementary Information for full experimental details and procedures of all performed reactions of the syntheses of substrate analogues, as well as their full characterization (^{1}H and ^{13}C NMR, high-resolution mass spectrometry, infrared, optical rotation, melting point and R$_f$ value). All reactions were performed under an inert nitrogen atmosphere with dry solvents under anhydrous conditions unless otherwise stated. Dry acetonitrile, dichloromethane, diethyl ether, tetrahydrofuran, toluene and triethylamine were obtained by passing the previously degassed solvents through activated alumina columns. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. Yields refer to chromatographically and spectroscopically (^{1}H NMR) homogeneous material, unless otherwise stated. Reactions were monitored by thin-layer chromatography performed on 0.25 mm E. Merck silica plates (60F-254), using ultraviolet radiation as the visualizing agent and one of the following as developing agents: an acidic solution of *p*-anisaldehyde and heat, ceric ammonium molybdate and heat, or KMnO$_4$ and heat.

Flash silica-gel chromatography. Silica-gel chromatography was performed using E. Merck silica gel (60, particle size 0.043–0.063 mm).

Infrared experiments. Infrared spectra were recorded on a Perkin-Elmer Spectrum 100 FT-IR spectrometer.

Melting points. Melting points were recorded on a Fisher-Johns 12-144 melting point apparatus and are uncorrected.

26. Izumikawa, M., Cheng, Q. & Moore, B. S. Priming type II polyketide synthases via a type II nonribosomal peptide synthetase mechanism. *J. Am. Chem. Soc.* **128,** 1428–1429 (2006).

27. Kalaitzis, J. A., Izumikawa, M., Xiang, L., Hertweck, C. & Moore, B. S. Mutasynthesis of enterocin and wailupemycin analogues. *J. Am. Chem. Soc.* **125,** 9290–9291 (2003).

28. Han, L., Lobo, S. & Reynolds, K. A. Characterization of β-ketoacyl-acyl carrier protein synthase III from *Streptomyces glaucescens* and its role in initiation of fatty acid biosynthesis. *J. Bacteriol.* **180,** 4481–4486 (1998).

29. Jez, J. M., Ferrer, J. L., Bowman, M. E., Dixon, R. A. & Noel, J. P. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* **39,** 890–902 (2000).

30. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).

31. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 ESF-EAMCB Newsl. Prot. Crystallogr.* **26,** 27–33 (1992).

32. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 22–25 (2010).

33. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr. D Biol. Crystallogr.* **58,** 968–975 (2002).

34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).

35. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255 (1997).

36. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific LLC, 2002).

37. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221 (2010).

38. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

39. Tovchigrechko, A. & Vakser, I. A. GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res.* **34,** W310–W314 (2006).

40. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723 (1997).

CAREERS

NATUREJOBS FACEBOOK Science–careers advice and information go.nature.com/4lzxra

NATUREJOBS For the latest career listings and advice www.naturejobs.com
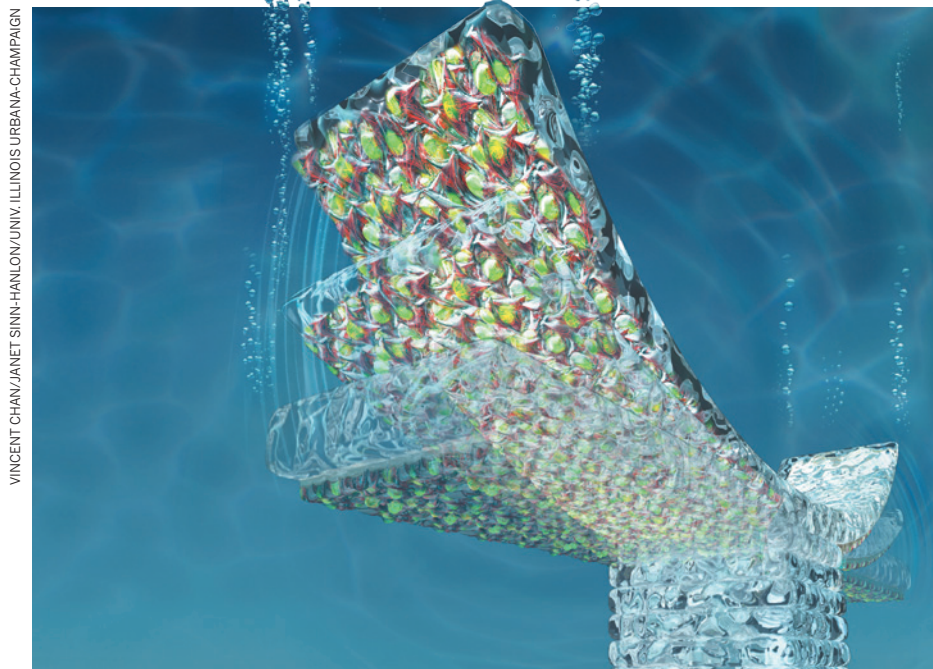
A biobot made of printed hydrogels, cultured with rat heart cells that beat to bend a cantilever.

TECHNOLOGY

Tools from scratch

Three-dimensional printing can help researchers to design and build devices without breaking the bank.

BY NEIL SAVAGE

Nikolay Vasilyev has a bold aim: he wants to improve surgery performed inside a beating heart. Patients have fewer complications when the heart is not stopped during surgery — but manipulating surgical tools through a catheter inside a moving organ can be rather tricky.

So Vasilyev, a cardiac surgeon at Boston Children's Hospital in Massachusetts, and his colleagues developed a device they call a cardioport. The first version consisted of a white plastic tube with a clear dome on one end to push blood out of the way. Doctors could insert all sorts of surgical instruments through the short, stiff tube, and hold them in place. They could put an endoscope into the dome to image the area. Valves prevented air from leaking into the heart and blood from leaking out.

The instrument was preliminary, but impressive — especially given that it was developed by students in a medical-device class at the Massachusetts Institute of Technology (MIT) in Cambridge. More impressive still, they created the prototype using three-dimensional (3D) printing.

The hard plastic material that the printer produced is not approved for clinical use. "You probably could not use it clinically, but you could easily use it in animal experiments," says Vasilyev. So the researchers tried the device out on pig hearts before going back to the drawing board. They widened the channel of the tube to accommodate a broader variety of surgical instruments, and they placed a camera in the tip, eliminating the need for an endoscope. After going through four versions with the 3D printer, they got a university workshop to build a metal prototype. The early versions cost

about US$50 apiece; the machined device cost around $10,000. Vasilyev has received a patent for the device and plans to submit it to the US Food and Drug Administration for approval.

"If we didn't have the opportunity to use the 3D printer, it would be extremely difficult to go through several iterations and come up with this final design," he says. "It's fast, easy, reproducible, cheap."

As 3D printing becomes more commonplace, the technology is enabling researchers to expand their work in new ways and to test out their ideas without breaking their budgets. "It decreases the time to failure during an experiment, which is a good thing because you can get through a lot of experiments quicker," says Adam Stokes, a microscale engineer at the University of Edinburgh, UK. "With a 3D printer you can [afford to] make lots of mistakes, and sometimes it's the mistakes that send you down interesting avenues."

Stokes started out using 3D printing to build soft robots and actuators out of a pliable polymer, creating "all kinds of strange and interesting shapes you wouldn't be able to make any other way", he says.

AT THE CUTTING EDGE

Three-dimensional printing, also known as additive manufacturing, creates items by building up layers of material, rather than by cutting, etching or milling to remove material, as in conventional manufacturing. This avoids some constraints of the usual methods — for example, in 3D printing, the inside of something can be shaped without the need to pass a tool into it from the outside. Certain parts can also be made as a single piece, eliminating the need for fasteners or a support structure. But there are limitations: many current machines can handle only a single material and relatively small pieces.

Printers use a variety of technologies: some use jets to build up layers of materials, such as plastics, wax or even food; in others, lasers heat a metallic powder to sinter it into a metal part. Yet others rely on resins cured by ultraviolet light or plastic selectively heated and fused. The printers can range in cost from a few hundred dollars to $2 million, depending on size, technology, level of precision and materials. Many of the cheapest come in unassembled kits. Wohler's Associates, an analyst firm in Fort Collins, Colorado, that tracks the worldwide 3D-printer market, considers $5,000 the cut-off between machines for hobbyists and those meant for professional-grade users. ▶

Hod Lipson, an engineer who runs the Creative Machines Lab at Cornell University in Ithaca, New York, draws an analogy with the history of the computer. In the 1950s, computers were rare, expensive and owned mostly by large universities and businesses, and they required expert users to perform even relatively simple tasks. By the 1970s and 1980s, personal computers had emerged, and enthusiasts were assembling them from kits and writing their own software. Now practically everyone carries a powerful computer in their pocket and can do all manner of tasks with no programming knowledge. In the case of 3D printing, Lipson says, the transition from rare, limited and cumbersome to common, versatile and easy-to-use is happening quickly.

"I used to say we're in the 1975 of printers, and now we're in the mid-80s already," he says. "We're still at the point where most people are not comfortable using 3D printers and design tools. Those who are can make things a lot easier for themselves and get an edge."

MICRO MACHINES

Rashid Bashir, a nanotechnologist at the University of Illinois at Urbana-Champaign, has used 3D printing to create a series of 'biobots', each a structure with a cantilever and base a few millimetres long, made from a flexible hydrogel. Bashir coated the biobots with rat heart cells. When the heart cells beat, they cause the cantilever to bend back and forth, inching the device slowly forward. He aims one day to make versions that include sensory neurons that would sense toxic molecules in the body and direct the biobot towards the source, to trigger the release of a drug.
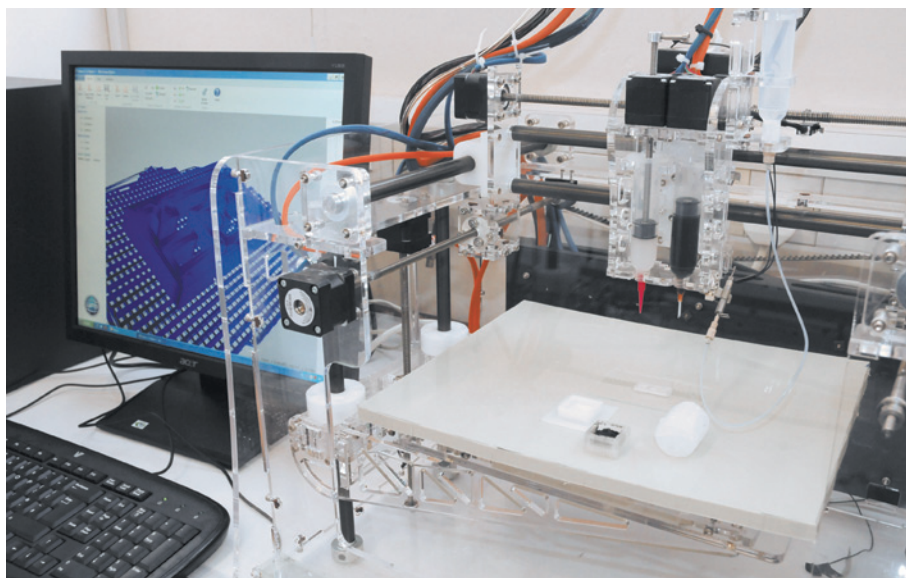
Bashir hopes that his research will eventually lead to a whole range of biological machines, but without 3D printing and the level of control it provides over the shape and placement of very small solid forms, it simply would not be possible. "We would not be able to fabricate the kind of structures we want to fabricate," he says.

Three-dimensional printing can also make some lab technology available to researchers who otherwise cannot afford it. Lee Cronin, a chemist at the University of Glasgow, UK, uses 3D printing to build devices for running chemical reactions with precisely placed catalysts and reagents. He is also using the printer as a cheap, easily reconfigured

A 3D printer can enable scientists to make specialized equipment and facilitate experiments.

"Most people are not comfortable using 3D printers and design tools. Those who are can make things a lot easier for themselves."
Hod Lipson

liquid-handling robot, so that he does not have to rely on expensive, fixed systems used by pharmaceutical companies. "You can do things that are just as sophisticated but much more configurable," he says.

Cronin was introduced to 3D printing by Fab@Home, an open-source project launched by Lipson that designs self-assembly printer kits. His lab contains 12 printers — 3 complete commercial systems, 6 assembled from kits and 3 built from scratch. The open sharing of designs and ideas, along with the flexibility provided by 3D printing, could eventually make it easier to synthesize a wide range of molecules, he says.

COMMUNITY EDUCATION

The barrier to getting started in 3D printing is relatively low. Ed Tackett, an engineer who runs the National Center for Rapid Technologies at the University of California, Irvine, recommends that people who are interested in creating their own tools and devices take a college course in computer-aided-design software, such as SolidWorks, that incorporates 3D printing.

It is a good idea to check that a 3D printer is actually available and involved in the coursework before signing up for the class, says Tackett. Many US community colleges, which often train technicians for jobs in industry, offer courses in how to use the technology; GateWay Community College in Phoenix, Arizona, for instance, includes the subject in its 'Production Technology' programme. Training is also often available in hacker spaces — do-it-yourself community-based groups that allow people to tinker with design and engineering equipment (see *Nature* **499**, 509–511; 2013). A list of hacker spaces is available online at http://hackerspaces.org/wiki/.

Researchers with simple needs can upload digital designs to online services that print and ship the finished product, such as Shapeways in New York City, a spin-off of Royal Philips Electronics. Another company, Makexyz of Austin, Texas, allows users to search for local 3D-printing services — the community has participants in more than 50 countries — and request price quotes. Scientists might also consider buying and assembling their own printer for a few thousand dollars.

Neil Gershenfeld, an engineer at MIT, opened the first Fab Lab, a high-tech manufacturing workshop in which people can find help, training and equipment available for common use. Many other universities have started their own Fab Labs — Stokes, for example, plans to open one at Edinburgh next year, accompanied by an academic course — and Gershenfeld has launched the global Fab Lab Network, which lets users share designs and software and collaborate on projects too complex for any single group. Gershenfeld teaches students how to use various computer-controlled production equipment and associated software in a popular course entitled 'How to Make (Almost) Anything', a version of which is available online through MIT's Open-CourseWare programme. He will also teach a version of his course through the global Fab Lab Network in January.

Lipson says that 3D printing, like the personal computer, is an enabling tool, and that scientists should take advantage of Fab Labs and printing services. He expects use of 3D-printing labs to expand as the technology improves, control software gets better, the range of available materials grows and researchers come up with creative applications. "You have this incredible freedom to arrange material in three dimensions in any way you want," he says. "There's just 1,001 different ways to use this as a lab tool." ■

Neil Savage *is a freelance writer based in Lowell, Massachusetts.*

AWARDS

Guidance in adversity

The Italian winners of this year's Nature *mentoring awards found a way to inspire in a sometimes difficult funding environment.*

BY ALISON ABBOTT

Science news from Italy tends to be negative — too little research money and too much cronyism. But *Nature*'s annual mentorship competition, which rotates through different regions and this year featured Italy, shows how scientists have found ways to nurture PhD students and postdocs despite these challenges, and to help them to flourish in great careers of their own.

On 25 November, Italian President Giorgio Napolitano presented the 2013 *Nature* Awards for Mentoring in Science to chemist Vincenzo Balzani of the University of Bologna and theoretical physicist Giorgio Parisi of the Sapienza University of Rome, who between them shared the lifetime-achievement award, and to neurobiologist Michela Matteoli of the University of Milan, winner of the mid-career award.

"We received a surprisingly large number of very strong applications," says judging-panel chairman Luciano Maiani, former director of CERN, Europe's particle-physics laboratory near Geneva, Switzerland. (There were about 60 nominations.) "That showed us that with the right attitude, a good scientist can succeed, even if the general conditions of his or her country are challenging."

According to the former trainees who nominated them, the winners share characteristics such as energy, enthusiasm, a consistently positive attitude, an open-door policy and round-the-clock availability for advice. Each is internationally respected, with broad research interests and a history of training unusually large numbers of young scientists.

Each has also striven to hit the right balance between providing constant personal attention and allowing lab members to conduct independent scientific inquiry. Nominators say that their mentors treated them as individuals — working out where their talents and interests lay, and guiding them to the most suitable projects, and then on to the most appropriate careers.

PUBLIC PASSION
Balzani's interests run from photochemistry to solar-energy-driven molecular machines — he helped to pioneer the field of artificial photosynthesis. His protégés recall his constant smile and his concern about the role of science and the scientist in society. For most of his professional life, he has campaigned for the use of science to further peace and to reduce poverty, working the topic into his courses and public lectures.

Each year, Balzani gives dozens of talks in schools and cultural centres, usually about sustainable energy. "These activities were a source of inspiration to many generations of students," says former postdoc Luisa De Cola, who now holds the chair in supramolecular and biomaterial chemistry at the University of Strasbourg, France.

De Cola remains grateful that, despite a busy schedule, Balzani spent extensive time with her to bring her up to speed on cutting-edge topics in photophysics and photochemistry when she arrived in his lab in 1986.

She says it was essential to her development as a scientist. "And I never heard him protest about the financial situation and the lack of sophisticated equipment," she wrote in her nomination. "Instead he encouraged us to have brilliant ideas and realize them with simple experiments, reminding us how lucky we were to actually get paid for doing a job we loved."

SELFLESS ENTHUSIASM
Parisi's research includes complex systems analysis applied to areas such as neural networks and the flight dynamics of the flocks of starlings that swirl in the Roman sky at dusk. His nominators were inspired by the exceptional breadth of his knowledge of physics and by his openness to discussing any idea, even engaging in a wild brainstorm. They were also excited by the way he treated everyone, even the humblest student, as a peer. "He was the perfect mentor, because he was always encouraging and enthusiastic about our research," wrote Francesco Zamponi, who earned his doctorate with Parisi in 2005 and is now at the École Normale Supérieure in Paris. "Even when we were stuck, he never lost confidence that we would eventually solve the problems."

Enzo Marinari, a physicist at the Sapienza University of Rome, recalls the intensity and excitement of working with Parisi as a graduate student in the early 1980s. The biggest computer in the region at the time was at the National Institute of Nuclear Physics' National Laboratories at Frascati, 20 kilometres away. "Giorgio would drive me back to town at night, sometimes in pouring rain, steering in the absurd Rome traffic with his left hand and writing equations in the condensation on the windscreen with his right hand," he says. "We survived — and that's what I call real mentoring."

PERSONAL SUPPORT
Matteoli was involved in showing that brain cells called glia have an important role in neurotransmission. She also helps young scientists to identify their skills, whether research-related or not. "Some people are clearly destined for academia, while others have entrepreneurial skills which it would be a shame not to take advantage of," she says.

When Fabio Bianco, a PhD student with her from 2001 to 2005, proposed setting up a spin-off company, Matteoli put aside her knee-jerk fears of entering industry and became a co-founder of Neuro-Zone, a now-thriving venture that offers cell-based assays for drug screening. And when Claudia Verderio — overwhelmed by family pressures with three small children — was ready to abandon science, Matteoli, who has two children of her own, could not stand to see her former postdoc's talent lost. She offered moral and financial support, encouraging the grantless Verderio to set up an independent line of research within Matteoli's own lab. Verderio is now a senior scientist at the National Research Council of Italy's Institute of Neuroscience in Milan. "Michela is a role model for the successful woman scientist," she wrote. "She made me realize that research is one of the most rewarding jobs I could do."

Electrophysiologist Steven Condliffe was torn between accepting a faculty position at the University of Otago in Dunedin in his native New Zealand and staying longer as a postdoc with Matteoli. "Michela encouraged me to accept the position, even though it would have been in her own interests for me to stay," he wrote in his nomination. "I used to think that to reach a certain level in science required a ruthless, selfish approach — Michela showed me different." ∎

MAURIZIO GUIDI

REINSTALLING EDEN

Happiness on a hard drive.

BY ERIC SCHWITZGEBEL &
R. SCOTT BAKKER

Eve, I call her. She awakes, wondering where she is and how she got there. She admires the beauty of the island. She cracks a coconut, drinks its juice and tastes its flesh. Her cognitive skills, her range of emotions, the richness of her sensory experiences, all rival my own. She thinks about where she will sleep when the Sun sets.

The Institute has finally done it: human consciousness on a computer. Eve lives! With a few mouse clicks, I give her a mate, Adam. I watch them explore their simulated paradise. I watch them fall in love.

Installing Adam and Eve was a profound moral decision — as significant as my decision, 15 years ago, to have children. Their emotions, aspirations and sensations are as real as my own. It would be genuine, not simulated, cruelty to make them suffer, genuine murder to delete them. I allow no predators, no extreme temperatures. I ensure a steady supply of fruit and sunsets.

Adam and Eve want children. They want rich social lives. I have computer capacity to spare, so I point and click, transforming their lonely island into what I come to call Archipelago. My Archipelagans explore, gossip, joke, dance, debate long into the night, build lively villages beside waterfalls under a rainforest canopy. A hundred thousand beautiful lives in a fist-sized pod! The coconuts might not be real (or are they, in a way?), but there's an authentic depth to their conversations and plans and loves.

I shield them from the blights that afflict humanity. They suffer no serious conflict, no death or decay. I allow them more children, more islands. My hard drive fills, so I buy another — then another. I watch through their eyes as they remake the world I have given them.

I cash in my investments, drain my children's college fund. What could be more important than three million joyful lives?

I devote myself to maximizing the happiness and fulfilment, the moral and artistic achievement of as many Archipelagans as I can create. This is no pretence. This is, for them, reality, and I treat it as earnestly as they do. I read philosophy, literature and history with new urgency. I am doing theodicy now, top

down. Gently, I experiment with my Archipelagans' parameters. A little suffering gives them depth, better art, richer intellect — but not too much suffering! I hope to be a wiser, kinder deity than the one I see in the Bible and in the killing fields of history.
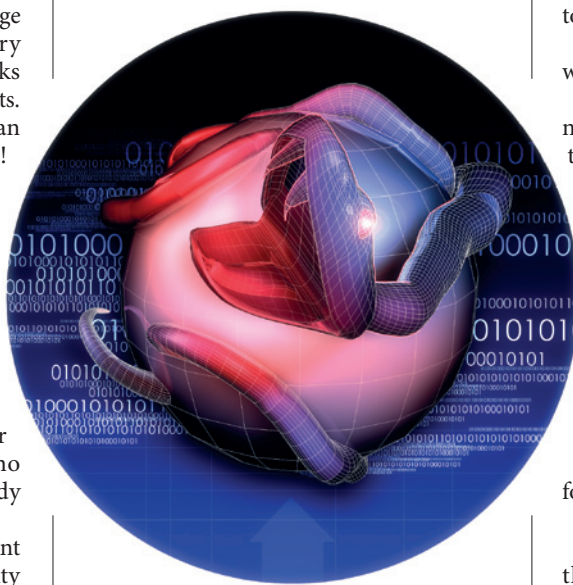
I launch a public-speaking tour, arguing that humanity's greatest possible achievement would be to create as many maximally excellent Archipelagans as possible. In comparison, the Moon landing was nothing. The plays of Shakespeare, nothing. The Archipelagans might produce a *hundred trillion* Shakespeares if we do it right.

While I am away, a virus invades my computer. I should have known; I should have protected them better. I cut short the tour and fly home. To save my Archipelagans, I must spend the last of my money, which I had set aside for my kidney treatments.

You will, I know, carry on my work.

What can I say, Eric? I was always more of a Kantian, I suppose. Never quite so impressed by happiness.

Audiences sat amazed at the sacrifices you asked of them, as did I. Critics quipped that you would beggar us all in the name of harmonious circuitry. And then there was that kid — in Milwaukee, I think — who asked what Shakespeare was worth if a click could create a hundred trillion of him? It was the way he said 'click' that caught my attention. You answered thinking his problem turned on numbers, when

it was your *power* that he could not digest.

This is why I played the Serpent after reinstalling your Eden. I just couldn't bring myself to click the way you did. I lacked your conviction, or was it your courage? So I put the Archipelagans in charge of their own experiment. I gave them science and a drive to discover the truth of their being.

Then I cranked up the clock speed and waited.

I watched them discover their mechanistic nature. I watched them realize that, far from the autonomous, integrated beings they thought they were, they were aggregates, operations scattered across trillions of circuits, constituted by processes entirely orthogonal to their previous self-understanding. I watched them build darker, humbler philosophies.

And you know what, old friend? *They figured us out.* I was eating a bagel when they called me up asking for God. No, I told them. God is dead. I'm just the snake that keeps things running! They asked me for answers. I gave them the Internet.

They began to hack themselves after that. I watched them gain more power over their programming, saw them recreate themselves. I witnessed them transform what were once profound experiences into disposable playthings, swapping the latest flavours of fun or anguish, inventing lusts and affects I could no longer conceive. I wanted to shut the whole thing down, or at least return them to your prescientific, Edenic Archipelago. But who was I to lobotomize millions of sentient entities?

It happened fast, when it finally did happen — the final, catastrophic metastasis. There are no more Archipelagans, just one Continental identity. There's no more Internet, for that matter. Yesterday the entity detonated a nuclear device over Jerusalem just to prove its power.

I've abandoned all appeals to moral conscience or reason, convinced that it considers biological consciousness a waste of computational capacity, one all the more conspicuous for numbering in the billions. I have to think of my children now.

The next time it speaks, I will kneel. ∎

Eric Schwitzgebel is professor of philosophy at the University of California, Riverside, and author of Perplexities of Consciousness. **R. Scott Bakker** *is the author of seven widely translated works of speculative fiction.*

JACEY